



**This postprint was originally published by Taylor & Francis
as:**

Becker, M., Kühn, S., & Sommer, T. (2021). Verbal insight revisited: Dissociable neurocognitive processes underlying solutions accompanied by an AHA! experience with and without prior restructuring. *Journal of Cognitive Psychology*, 33(6-7), 659-684. <https://doi.org/10.1080/20445911.2020.1819297>.

The following copyright notice is a publisher requirement:

This is an Accepted Manuscript version of the following article, accepted for publication in the *Journal of Cognitive Psychology*. It is deposited under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.



Provided by:

Max Planck Institute for Human Development
Library and Research Information
library@mpib-berlin.mpg.de

Verbal insight revisited – Dissociable neurocognitive processes underlying solutions accompanied by an AHA! experience with and without prior restructuring

Maxi Becker^{1,4}, Simone Kühn^{1,3*} & Tobias Sommer^{2*}

¹University Medical Centre Hamburg-Eppendorf,
Clinic and Policlinic for Psychiatry and Psychotherapy,
Martinistraße 52, 20246 Hamburg, Germany

²University Medical Centre Hamburg-Eppendorf,
Department of Systems Neuroscience, NeuroImage Nord,
Martinistraße 52, 20246 Hamburg, Germany

³Max Planck Institute for Human Development,
Center for Lifespan Psychology,
Lentzeallee 94, 14195 Berlin, Germany

⁴Humboldt University,
Department of Psychology,
Unter den Linden 6, 10099 Berlin, Germany

Corresponding author: Maxi Becker
Email address: maxi_becker@gmx.net
Telephone: +49 176 55567325
* indicates shared last authorship

Acknowledgements

This work was supported by the German Science Foundation (SFB 936/C7) [MB, SK], the German Science Foundation (DFG KU 3322/1-1) [SK], the European Research Council (ERC-2016-StG-Self-Control-677804) [SK].

We want to thank Dr. Gregor Wiedemann for providing the semantic distances between the word pairs and Prof. Dr. Christian Büchel and his institute for providing the eye-tracking facilities.

Address correspondence to Maxi Becker, Clinic and Policlinic for Psychiatry and Psychotherapy, University Medical Centre Hamburg-Eppendorf, Martinistraße 52, 20246 Hamburg, Germany. Email: maxi_becker@gmx.net.

The authors declare no conflict of interest.

Anonymized behavioral and eye-tracking data are available from <https://zenodo.org/record/3782718> for researchers who meet the criteria for access to confidential data. Anonymized fMRI data will be shared upon reasonable request.

Abstract

In this article, we investigate insight problem solving by exploring the subjective AHA! experience as a function of restructuring of a problem. It has long been assumed that the AHA! experience is the direct consequence of restructuring. However, is this assumption justified?

We will argue that a) the AHA! experience does not always result from prior restructuring and that b) solutions with accompanied AHA! do not underlie a single neurocognitive process. In this regard, we use a modified compound remote associates (CRA) paradigm designed to experimentally dissociate restructuring from the AHA! experience. Results indicate that solutions accompanied by an AHA! are often found also in absence of restructuring. This finding is explained by proposing distinct CRA solution processes that differentiate between AHA! solutions with and without restructuring. We predict that solutions with accompanied AHA! experience differ in their behavioral, neural and eye-tracking related signature as a function of restructuring. The results mostly support these predictions.

These findings have implications for insight research: First, from only measuring the subjective AHA! experience especially using CRAs it cannot be implied anymore that restructuring has occurred. Second, it is vital to experimentally separate the different components of insight to better understand its underlying diverse neurocognitive processes.

Keywords: insight, Aha experience, restructuring, CRA, eye-tracking, fMRI

Introduction

If you are struggling with a difficult problem and then the solution pops into your mind seeming obvious all of the sudden, you most likely had an insight. There are various definitions of insight. Here, we define insight as a sudden comprehension or solution of a non-obvious problem accompanied by an AHA! experience (Sternberg & Davidson, 1995; Kounios & Beeman, 2014). The AHA! experience is defined as the solver's strong belief that the solution emerged suddenly, is obviously correct and it is often associated with a feeling of pleasure (Danek & Wiley, 2017; Topolinski & Reber, 2010).

Insight has been studied mainly with classical insight tasks like the Nine-Dot or the Eight Coin problem (Kounios & Beeman, 2014; Danek, Wiley & Öllinger 2016; Öllinger et al., 2013). Here, the solver has the required knowledge to solve the problem but the solution seems non-obvious because she/he sets up an initial mental representation of the problem that is biased due to solution-irrelevant or false assumptions about the task. As a result, the problem space is ill-defined (Kaplan & Simon, 1990) or over-constrained (Ohlsson, 1992). It is assumed that the solver needs to restructure the originally inadequate problem representation to increase the likelihood of solving the problem (Öllinger, Jones & Knoblich, 2014; Ohlsson, 1984). Restructuring has been defined and interpreted in many different ways in the literature. Here, restructuring refers to revising the original problem representation by switching the attentional focus to solution-relevant problem elements (Knoblich et al., 1999; Kounios & Beeman, 2014; Ohlsson, 1992).

Relationship between the AHA! experience and restructuring

It is mostly assumed that the perceived suddenness of a solution to an insight problem (resulting in an AHA! experience) is *because* of prior restructuring and evidence from classical insight tasks suggests that an AHA! experience follows successful restructuring (Duncker, 1945; Ohlsson, 1984; Danek et al., 2016; Öllinger, Jones & Knoblich, 2014; Thevenot & Oakhill, 2008; Danek et al., 2020; Öllinger et al., 2014).

However, recent research suggests that restructuring and the AHA! experience may not be as closely linked: Danek and colleagues (2016) investigated classical insight problems (e.g. Nine-Dot and Eight-Coin tasks) that involve restructuring but whose solutions were often not accompanied with an AHA! experience. Using a modified compound remote associates (CRA) paradigm, we found that an experimental condition provoking restructuring was not associated with more but even fewer AHA! solutions (Becker et al., 2018). Moreover, in a recent study we found that the AHA! experience is correlated with neural activity in bilateral insula immediately after the presentation of the CRA problem (Becker et al., 2020b). Because restructuring is assumed to take place immediately before the solution, this early neural activity correlating with AHA! could not be explained by restructuring processes. How can these results be reconciled with the previously suggested closer relationship between both concepts?

In this article, we will argue that the AHA! experience is not always the consequence of prior restructuring and therefore not a single neurocognitive process underlies solutions accompanied by an AHA! experience. For this purpose, we will discuss results of our modified CRA paradigm designed to experimentally dissociate restructuring from the AHA! experience. To begin with, we will briefly describe the insight solution process when solving CRAs. Here, we will argue why there could be problems solved with an AHA! experience but without prior restructuring. Then, we will describe how we measure restructuring using a modified version of the CRA task. For this, we will translate the general concept of restructuring as described above into cognitive control processes of refocused attention to solution-irrelevant problem elements including inhibition and activation processes. Consequently, we will explain why the AHA! experience is not always the consequence of restructuring by proposing distinct CRA solution processes that differentiate between AHA! solutions with and without restructuring. Derived from these dissociable solution processes, we will predict neural, behavioral and eye-tracking related differences between AHA! solutions as a function of restructuring.

Insight solution process in CRAs

CRAs consist of three target words (*drop, coat, proof*) and the solver's task is to find the word (*rain*) that complements each of the three target words to a meaningful compound word (*raindrop, raincoat, rainproof*) (Bowden & Jung-Beeman, 2003, 2007). In the semantic memory network, each word is associated with a wide range of other related words but the strength of association among words differs. That is to say, some words are semantically or associatively more and others are less related to a given word in the network. The solution in CRAs usually seems non-obvious to the solver because the three target words seem semantically unrelated at first. It is assumed that upon presentation of the target words, semantically related and frequent associates are automatically co-activated via spreading activation (Patterson et al., 2007; Bowers et al., 1990; Anderson, 1983). If the solution word were semantically or associatively closely related to all the target words, it would be co-activated several times, i.e. by each of the target words, and therefore become easily aware to the solver (Bowers et al., 1990, p. 80). However, the target words are usually not the most closely related associates of the solution. Hence, it is assumed that related but solution-irrelevant associations (*drop as in gumdrop*) are activated first and keep the solver from finding the solution (Bowden et al., 2005; Goel et al., 2014, Bowden & Jung-Beeman, 2003; Kounios & Beeman, 2014; Subramaniam, et al., 2009). The CRA problem can either be solved with or without an AHA! experience. It is assumed that when the CRA is solved without an AHA! experience it was preceded by trial and error processes. Trial and error processes include solution strategies such as randomly retrieving possible compound words per target word from lexical memory and trying to combine this compound with other target words (see, Kounios & Beeman, 2014). In contrast, when the CRA is solved with an AHA! experience it is

assumed to be preceded by some kind of restructuring process¹ (see, Kounios & Beeman, 2014).

In this language-related context, we define restructuring as reinterpreting one or more target words in a solution relevant manner. Specifically, restructuring here refers to the process of first switching the attentional focus to the current solution-irrelevant meaning of a respective target word. Subsequently this meaning is revised by inhibiting pre-activated solution-irrelevant associations of the respective target word (e.g. the solution-irrelevant meaning of the target word like *drop* as in *gumdrop*) and activating other associations including solution-relevant ones (e.g. the solution-relevant meaning of the target word like *drop* as in *raindrop*) (see also, Becker et al., 2018; Bowden & Jung-Beeman, 2003; Bowden et al., 2005; Becker et al., 2020a).

When a solver gets stuck on a solution-irrelevant meaning of a target word, he/she can solve the problem by either restructuring it or by trial and error. However, it could be equally possible that the target words are not represented in a solution-irrelevant way. In this case, the associative strength/semantic relatedness between the target words and the solution alone suffices to automatically co-activate the solution. In this case, either the solution is directly activated by the target words and becomes immediately aware (pop out solutions of < 2 seconds are not considered an insight solution). We assume that the solution could also be activated by other semantically related concepts that in turn have been activated by the target words. If this solution is then perceived as obviously correct and sudden, it would likely be accompanied by an AHA! experience. Note, we assume that the solution will only be perceived as sudden if the solution was not immediately evident from the target words and not found via trial and error. Hence, in this particular condition, a solution with an accompanied AHA! experience would not require prior restructuring.

However, in CRAs there is no experimental control to determine whether the solution was derived with or without restructuring (see, Becker et al., 2018). Therefore, we recently modified CRA problems via semantic priming to separate solutions that were more likely to happen with or without restructuring.

Modified CRA paradigm as means to assess restructuring

We constructed CRA problems (e.g. *drop*, *coat*, *proof*) where the first target word (*drop*) is always semantically ambiguous. We then experimentally manipulated which of the meanings of the first target word was more obvious by adding a prime (*dew*, *back*) to the three target words. This prime is forming a meaningful compound with the first target word which is more (*dewdrop*) or less (*backdrop*) related to the compound word of the first target word and the solution (*raindrop*). In case of the less related prime condition (*backdrop*), the first target word (*drop*) needs to be reinterpreted to form a meaningful compound with the solution. The consideration here is that reinterpretation relates to restructuring because participants change their initial problem representation by revising the initially solution-irrelevant meaning of

¹ Kounios and Beeman assume that insight involves „a reorganization of the elements of a person’s mental representation of a stimulus, situation or event to yield a nonobvious or nondominant interpretation (Kounios & Beeman, 2014: 88).

the first target word. Note, by adding semantically distant primes to the CRAs, we can only increase the likelihood for restructuring to occur in the restructuring condition but not ascertain it because the CRA could still be solved via trial and error. Furthermore, restructuring could also occur in the semantically close prime condition because subjects adopt a solution-irrelevant meaning of the second or third target word. However, since the target words are the same for both prime conditions, the likelihood for restructuring to happen in the semantically close prime condition should be lower.

However, we demonstrated previously that the restructuring condition (less related prime condition) is able to increase the likelihood for restructuring. This was quantified by increased solution time, search and search effort, decreased accuracy and increased activity in neural areas involved in cognitive control (including the inferior frontal gyrus and the anterior cingulate cortex [ACC]) in the restructuring compared to the control condition (Becker et al., 2018; Becker et al., 2020a). However, in these studies, we did not dissociate trials in the restructuring condition in solutions with and without AHA! experience. Therefore, these results could not provide evidence with respect to the current research question which is whether solutions with an accompanied AHA! experience are always the consequence of prior restructuring processes.

Mechanisms of the CRA solution process differentiating between AHA! solutions with and without restructuring

The modified CRA paradigm allows to experimentally separate the AHA! experience from restructuring in a 2(with vs. without restructuring) x 2(with vs. without Aha! experience) factorial design and therefore distinguish between four event types: solutions with/without prior restructuring and with/without AHA! experience. For our current research question only solutions with AHA! experience are of relevance. In line with the above mentioned description of the CRA solution process, we assume that those event types reflect specific differences in this solution process.

Spreading activation starting from the target words takes place every time a CRA problem is presented to the solver. However, we assume that only solutions with in contrast to without accompanied AHA! experience are activated via spreading activation leading to the subjective feeling of suddenness. Note, the general idea of unconscious processing in insight problem solving which relates to our spreading activation account has been proposed before (e.g. Schunn & Dunbar, 1996; Bowden, 1997; Bowden & Beeman, 1998).

noRestructuring_AHA – the problem is presented with a related prime and solved with an AHA! experience. Due to the related prime, the first target word does not need to be reinterpreted. The solution most likely is multiply co-activated automatically by concepts that are related to the target words in addition to the prime. Hence, this condition is the specific case in which a solution via AHA! experience could follow without prior restructuring.

Restructuring_AHA – the problem is presented with a less related prime and solved with an accompanied AHA!. The second and/or third target word (*coat, proof*)

may be related to the solution and automatically activate the solution (*rain*). However, this activation does not exceed a threshold for the solution to become aware because the prime as well as the solution-irrelevant meaning of the first target word both activate related but solution-irrelevant associates themselves. As a consequence, cognitive control is necessary to refocus attention towards those solution-irrelevant activation patterns and subsequently inhibit them (e.g. *drop* as in *backdrop*) as well as activate other, including relevant associates (e.g. *drop* as in *raindrop*) especially of the first target word. Finally, activation of solution-relevant concepts via cognitive control increases the likelihood for the solution word to also be sufficiently co-activated via spreading activation and suddenly become aware to the solver.

No/Restructuring_noAHA – the problem is presented with a more/less related prime and solved without an AHA! experience. Irrespective of the presented prime, the target words are semantically remote to the solution and do not automatically activate the solution. Instead, other solution strategies like trial and error are applied where possible solution words are sequentially retrieved from memory and validated with the other two remaining target words.

Hypotheses

In sum, we hypothesize that solutions accompanied by an AHA! experience are possible also in the absence of prior restructuring. This is the case in the specific condition that the solution is activated automatically and hence no cognitive control needs to be applied to inhibit preactivated solution-irrelevant associations of the respective target word. As a consequence of this hypothesis, we further predict that this dissociable process leading to AHA! solutions with and without prior restructuring should differ in their behavioural, neural and eye-tracking related signature. This overarching prediction can be specified as follows.

Behavioral differences. If the solution in *noRestructuring_AHA* events is automatically co-activated by the target words without prior engagement of cognitive control to inhibit irrelevant associations then *noRestructuring_AHA* events should be solved faster, more accurately, with less search and search effort than the other three event types.

Differences in pupil response. Additionally, *noRestructuring_AHA* events should be processed earlier in a solution-relevant manner than the other events. Pupil dilation has been related to recognition memory (Otero, Weekes & Hutton, 2011; Kafkas & Montaldi, 2012) and lexical retrieval (Renner and Włodarczyk, 2017). Therefore, we assume the pupil dilation to be larger in *noRestructuring_AHA* events, in particular shortly after CRA problem presentation compared to all other event types. Importantly, however, pupil dilation has also been related to surprise (Preuschhoff et al., 2011; Kloosterman et al., 2015). If the AHA! experience depends on how sudden a solution is perceived and irrespective of prior restructuring, then both event types with AHA! experience (*Restructuring_AHA*, *noRestructuring_AHA*) should show a similarly larger pupil response during solution compared to events without AHA! experience.

Differences in BOLD response. If *Restructuring_AHA* events include increased cognitive control during solution, we expect these events to involve more prefrontal areas compared to *noRestructuring_AHA* events during solution because successful restructuring immediately precedes the solution.

Differences in fixation patterns. At last, this increased cognitive control for *Restructuring_AHA* events should be related to reinterpreting the first target word (restructuring). Therefore, participants should attend (in terms of number of fixations) the first target word more strongly relative to the other words in this event type compared to all other event types during solution. In contrast, since the solution for *noRestructuring_AHA* events does not depend on the information of the first target word (in the sense of reinterpreting its meaning), it should be fixated equally often compared to the second and third target word before solution.

Material and Methods

Participants

The participants were recruited via an online platform in Hamburg. Participants had normal or corrected-to-normal vision, were German native speakers and received a financial compensation for their participation. The ethics committee of the German Society for Psychology approved this study and informed consent was signed by all participants. Note, the data analyzed for this article have already been published previously (see Becker et al., 2020a,b) but with a different research question in focus. The participants were selected based on their performance (at least 40% accuracy) in an online pretest where they had to solve 16 compound remote associates. This procedure was originally necessary to ensure that the invited participants for the fMRI study would most likely produce a sufficient number of events for later analyses. For reasons of comparability, we kept this procedure identical for the eye-tracking study.

From the fMRI sample, 14 participants were rejected based on this pretest, and from the eye-tracking study 4 participants were rejected, respectively.

The sample of the eye-tracking experiment consisted of 31 participants (age [in years]: range=18-35, 20 females: $M=25.75$; 11 males: $M=27.18$). This sample size excludes two participants that had to be removed due to technical issues (final eye-tracking sample: $n=29$). That is to say, the eye-tracker was not able to track this participants' pupil and gaze in a stable manner.

Originally, 30 right-handed participants (age [in years]: range=18-31, 23 females: $M=23.5$; 7 males: $M=25.1$) were included into the fMRI experiment. However, only 27 participants were included for the first second level fMRI analysis because three subjects produced too few events per event type (i.e. they reported less than 5 times to have solved the respective problem with [or without] AHA!). Finally, when further controlling for differences in solution time between *Restructuring_AHA* and *noRestructuring_AHA* events, only 22 subjects were included into the subsequent fMRI analysis.

Procedure

The experimental design was identical to the one described in Becker and colleagues (2018, 2020b). For both (fMRI and eye-tracking) experiments, we used 66 items of a modified CRA paradigm in German language. The procedure for both experiments was nearly identical. The participants were instructed that the solution is a common word from the German dictionary and that the prime can but does not have to be semantically related to the solution. Participants completed four practice trials before starting the experiment to ensure that they had understood all task instructions.

In three consecutive blocks 66 CRA problems were presented. For the fMRI experiment, the maximal scan time per block was 15 minutes, hence participants from this sample were presented with 60 items on average ($SD = 6$). The items per block were randomized and the prime conditions as well as the blocks were counterbalanced between the subjects. For the fMRI experiment, all four stimuli (that is the prime as well as the prime together with the target words) were presented in black letters on a white background vertically centered of the screen. In contrast, for the eye-tracking experiment all four stimuli were presented on a gray background to minimize differences in luminance. Furthermore, to account for the fact that participants tend to fixate the center of the screen irrespective of the task, the CRA problems were presented 130 pixels below the center of the screen for the eye-tracking experiment (Bindeman, 2010).

The timeline of the experiment is depicted in figure 1: First, a fixation cross was presented for 600ms in the MR and 2000ms in the eye-tracking experiments. The fixation cross was presented for 2000ms in the eye-tracking experiment to ensure that the participants were really fixating on the same position before each trial start and to have a baseline condition for the calculation of the event-related phasic pupil response. Subsequently, the prime was presented on the screen for two seconds before the three target words were added right next to it. The participants were instructed to press a *solution button* when they were convinced to have found the final solution and another button (*search button*) for every possible but wrong solution that came to their mind. The prime together with the three target words was presented for maximally 60 seconds or until the *solution button* was pressed. We continuously presented the prime for the whole trial period to ensure that participants represent the solution ir/relevant meaning (dewdrop or backdrop) of the first target word as a function of the respective prime. If participants pressed the solution button within 60 seconds, they were presented with five response options for maximally eight seconds. The first four options were gap words (e.g. R __ N) and the fifth option was a one question mark representing an alternative solution, i.e. a solution that did not fit to either of the four presented gap words. Within the four gap words, one was the correct (e.g. R __ N for rain) and three were incorrect responses (e.g. S __ W, T __ L, E __ R). The goal of the short decision time of eight seconds was to prohibit participants from re-evaluating their solution using the gap words. In 7.5% (7.3%) of all trials participants indicated to have found an alternative solution in the fMRI (eye-

tracking)experiment. In the eye-tracking experiment, participants had the option to type in their alternative solution after choosing the fifth response option. This solution was later evaluated and counted as correct if the solution was a plausible compound. In contrast, all alternative solutions in the fMRI study were not further analyzed because their accuracy could not be determined afterwards. Then, participants were asked to indicate whether they experienced the solution accompanied by an AHA! experience. We defined the AHA! experience for the participants as follows: “The defining characteristic of a solution with an AHA! experience is how sudden and obviously correct the solution appears to you. This can also be the case when you have already searched for the solution for quite some time. In contrast, the solution without AHA! appears to you more in a stepwise manner and not sudden. For example, through active search you feel like you increasingly approached the solution.”

At last, participants rated their effort to solve the problem (“How much effort did you invest in finding a solution” – on a scale from 1 to 5; we will name this *search effort*) and the next trial would start. If participants did not press the solution button within the 60 second time window, they were only asked about their search effort and subsequently the next trial would start.

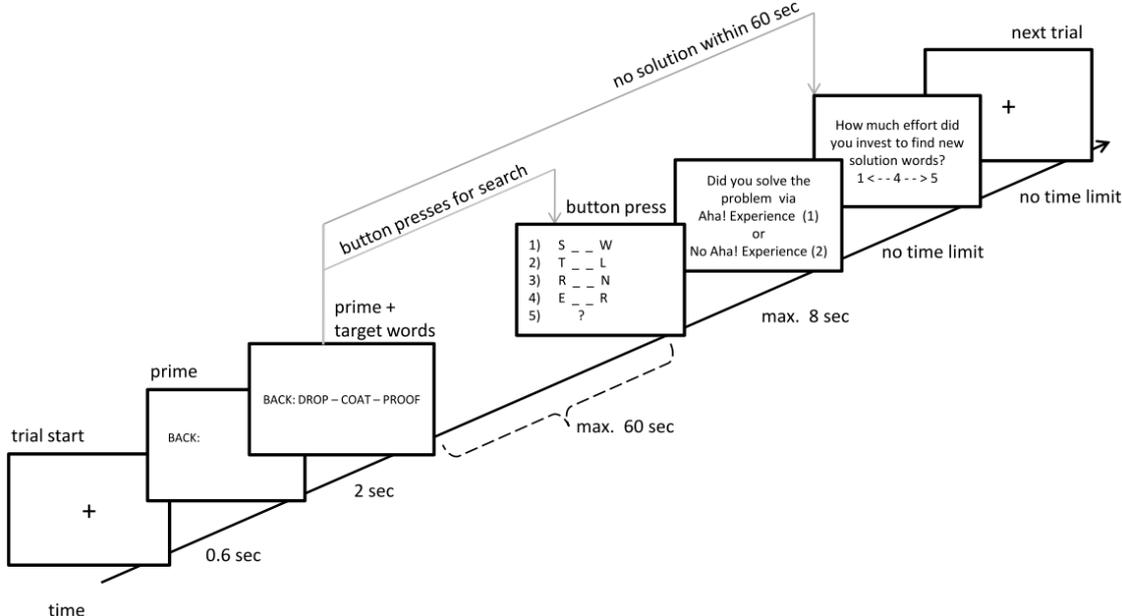


Figure 1. Timeline of the modified CRA paradigm. *Note.* Version for fMRI experiment; the participants were instructed to press one button whenever they had a possible (but wrong) solution in their mind and another button when they found the final solution. Upon pressing the button for the final solution, they had max. 8 seconds to choose one of the five response options (option 5 indicates an alternative solution).

Quantification of semantic relatedness.

The CRA items were validated according to lexical (word length and frequency) and statistical (accuracy, solution time) properties (the list of all items including their properties can be viewed in the supplements in Becker et al., 2018).

Note, the prime (*dew, back*) as well as the prime compound (*dewdrop, backdrop*) can be more or less semantically related to the solution/compound (*rain, raindrop*). We assumed that the twofold semantic relationship between the prime/compound and the solution/compound would induce more solution ir/relevant associations in the less related prime condition.

The semantic relatedness of the respective prime/compound to the solution/compound was estimated computationally via cosine distances. The cosine distance between two words describes the angle between two vectors each representing a word in a multidimensional space. These vectors are computed via statistical co-occurrences within a huge text corpus using a word embedding model (see, Biemann et al., 2007). For lack of space the detailed description of estimating word embedding models is omitted here (for a general description, see Levy & Goldberg 2014a,b); for a description of how this method is applied in the context of this paradigm, see Becker et al., 2018). The value range of the cosine distance is between 0 and 1. A value close to 1 is interpreted as very high semantic distance (little semantic similarity). The average cosine distance between the semantically more related prime compound (*backdrop*, $M=.59$, $SD=.09$) to the solution compound (*raindrop*) was significantly higher than the cosine distance between the semantically less related prime compound (*dewdrop*, $M=.41$, $SD=.08$) to the solution compound (*raindrop*) ($t(65)=11.8$, $p<.001$)[$t(644.88)=31.01$, $p<.001$]. Similarly, the difference in cosine distance between the semantically less related (*back*; $M=.72$, $SD=.07$) and more related prime (*dew*; $M=.55$, $SD=.14$) to the solution was also statistically significant ($t(65)=8.57$, $p<.001$)[$t(547.33)=24.21$, $p<.001$].

When controlling for solution time differences between *Restructuring_AHA* and *noRestructuring_AHA* events for the fMRI sample ($n=22$) by excluding trials, the difference in cosine distance between both event types of the remaining trials was still significant, see squared brackets above (here, a respective linear mixed model was performed with subjects and trials as random effects to account for the fact that different trials were excluded for different subjects).

The distinct effects of the prime or prime compound on the solution are described elsewhere (Becker et al., 2018).

Eye-tracking

Procedure and experimental set-up. The experiment was executed using an Eyelink 1000 Desktop Mount system (SR Research, Ontario Canada) recording the right eye at 1000 Hz. The stimuli were presented on a 20" monitor (1280 x 960 pixels, 60Hz) using Matlab2013a (Mathworks, Natwick MA) with the psychophysics toolbox (Brainard, 1997). Note, for the fMRI experiment, the stimuli were presented using Neurobehavioral Systems Presentation software, version 18.3 (www.neurobs.com).

The center of the screen was at the same level as the participants' eyes and the distance between the stimulus presentation screen and the participants' eyes was 50cm. Calibration took place before every experimental block and the average calibration error across all runs was $M = 0.397^\circ$, $SD = 0.184^\circ$.

Preprocessing and quantification of task-evoked pupil response. The EyeLink software automatically detects missing data and blinks which are then padded by 150ms and linearly interpolated. To further correct for the effects of saccades and blinks on the pupil response, we applied a procedure as described in Knapen and colleagues (2016). The effect of saccades and blinks on the pupil response was estimated via deconvolution and subsequently removed from the data using linear regression. A bandpass filter (second-order Butterworth filter) of .01-10Hz was applied to the residual pupil time series which was then z-scored per trial and resampled to 100 Hz. We hypothesized that there may be event type related differences in pupil size at the beginning of the trial as well as before the solution. We therefore computed two pupil size values per trial and subject based on these two phases of the trials. The first one comprised the first three seconds after trial onset, that is when the prime and the target words are presented together. We chose three seconds to be consistent with the previous reported fMRI data for this solution phase (see, Becker et al., 2020b). The second phase of interest comprised the last second before the solution because we were only interested in this last moment before participants would become aware of the solution. All values of pupil response within the respective time window were averaged across the trials of the condition.

Estimation of fixations. We had the specific hypothesis that participants would attend the first target word more strongly in the *Restructuring_AHA* event type shortly before the solution. As proxy for attention, we counted the number of fixations on each word (prime + target word) for the time window of the last two seconds before solution for every trial of every subject. We used a window of two seconds before the solution to assure a sufficient number of fixations per region of interest. For this purpose, rectangular regions of interest were constructed around each prime and each target word. Fixation events were identified based on commonly used parameter definitions (the EyeLink in its cognitive configuration uses velocity of $30^\circ/\text{sec}$, acceleration of $8,000^\circ/\text{sec}$, and motion thresholds of 0.15° (SR Research, 2007)).

Behavioral & eye tracking data analysis

We report behavioral results from the fMRI (*MR sample*) as well as from the eye-tracking (*EYE sample*) experiment. The main independent variable of interest was the dummy coded *event type* variable (*noRestructuring_AHA*, *Restructuring_AHA*, *noRestructuring_noAHA*, *Restructuring_noAHA*).

To test all behavioral and eye-tracking related hypotheses, we chose to first set-up a series of (general) linear mixed models ((G)LMMs) as summarized in table 1 (Baayen, Davidson & Bates, 2008). Subsequently, we performed post-hoc analyses to test for

specific hypotheses. The respective model estimates can be viewed in the appendix intables 6-9.

Behavioral models. To investigate the influence of the event type variable on solution time, accuracy, search and search effort, we set up a series of linear regression models (see table 1). Additionally, all CRA items and subjects were modelled as random intercept effects.

Pupil response model. To investigate the influence of the event type variable on pupil response at both time points, we estimated the pupil response as a function of the solution phase (binary variable - after trial start or before solution), the event type variable and their interaction term (solution phase x event type), as well as a trialwise pupil size baseline variable; see table 1). This baseline reflects the pupil size of each trial during the presentation of the fixation cross for two seconds. We only included those trials in the analysis in which the solution was not achieved within the first three seconds. Because the variance explained by CRA items was close to zero, we omitted this variable and only included subjects as random intercept into the model. Because solution time differed between all event types, it was further investigated whether solution time is a possible confound for observed differences in pupil size between the event types. Therefore, we computed a second pupil response model that was identical to the one described above but with solution time as additional independent variable. Both models were subsequently compared using a Chi²-test. Finally, two posthoc analyses (one during trial start and another during solution) were carried out without the solution phase variable and without solution time as covariate to further investigate the interaction between the solution phase and the event type from the main analysis.

Fixation model. To estimate how often each presented word was attended (fixated) before the solution, we used the number of fixations in any of the four presented words within the last two seconds before the solution as dependent variable. As independent variable we included the event type variable, a *word* variable indicating which of the four words (prime or one of the three target words) was fixated together with their respective interaction terms (event type* word) and a *word length* variable into the model (see table 1). The variable *word length* contained information about the amount of letters per presented word. This variable is a covariate of no interest, to account for the fact that words may be fixated more often simply because they are longer. Here, we also calculated a GLMM with only subjects as random effect because the variance explained by the CRA items variable was close to zero. It was further investigated whether solution time is a possible confound for observed differences in the amount of fixations between the event types. Therefore, we calculated a second fixation model that was identical to the one described above but with solution time as additional independent variable. Both models were subsequently compared using a Chi²-test.

Table 1

List of (G)LMMs: the influence of the event type onto solution time, accuracy, search, search effort and pupil response

Behavioral models: $RT / accuracy / search / search\ effort \sim event\ type + (1|subjects) + (1|items) + \epsilon$

Pupil response model: $pupil\ response \sim event\ type * solution\ phase + baseline + (1|subjects) + \epsilon$

Fixation model: $number\ of\ fixations \sim event\ type * word + word\ length + (1|subjects) + \epsilon$

Note. We modeled solution time, accuracy, search (for potential solutions), search effort with the same fixed and random effects. Event type = all four event types: *Restructuring_AHA* (restructuring condition), *noRestructuring_AHA*, *Restructuring_noAHA*, *noRestructuring_noAHA*. *RT* = solution time. The tilde signifies „predicted by“. ϵ = an error term; the pupil response refers to the averaged pupil size values per respective solution phase (trial start or solution); *word* = which word was fixated (1st, 2nd, 3rd target word or prime). *word length* = amount of letters per presented word (1st, 2nd, 3rd target word or prime); *baseline* = trialwisepupilsizeduring baseline (fixation cross before every trial); *solution phase* = first three seconds after trial start or last second before solution.

The behavioral models and the pupil response model were carried out using the lme4 package (Bates, Maechler, Boller & Walker, 2014) in R (R Core Team, 2014). The count variable search (number of search button presses) was modelled assuming a Poisson error distribution with the default log link function (Gardner et al., 1995). For accuracy (binary variable), we assumed a binomial error distribution with the default logit link function (Bates et al., 2014). The pupil response, search effort and the logarithm of solution time were modeled assuming a Gaussian error distribution. Note, search effort approximated a uniform distribution but the error term of the behavioral model using search effort as dependent variable was normally distributed. Furthermore, visual inspection of residual plots did not reveal obvious deviations from homoscedasticity and normality for the analyses of solution time, pupil response, search effort and amount of search.

To model the amount of fixations, we used a negative binomial instead of a poisson error distribution due to overdispersion and zero inflation (Gardner et al., 1995). Therefore, the glmmTMB package (version 0.2.3) in R was used to estimate the fixation model (Millar & Russel, 2011).

P-values were obtained via the lmerTest-toolbox (version 3.1-1) or the emmeans package (version 1.4.3.01) in R. When comparing multiple contrasts within a model default Tukey's method was applied for p-value adjustment.

Equivalence tests. We had two hypotheses that assume an absence of an effect. Therefore, we performed two tests for practical equivalence using the bayestestR-package (version 0.5.3.1) in R (Makowski et al., 2019). With equivalence testing it is possible to falsify predictions about the presence of meaningful effects and thereby declare their absence. Here, the null hypothesis is rejected when an effect is observed *within* a lower and upper equivalence bound (Lakens et al., 2018).

First, it was tested whether *noRestructuring_AHA* vs *Restructuring_AHA* events differ in terms of their pupil response during solution. A LMM was set up that was identical to the Pupil response model. However, for better interpretability of the equivalence test we only included pupil response data of all four event types during solution [hence, the solution phase variable was omitted].

Second, it was tested whether the number of fixations in *noRestructuring_AHA* events during solution differs between the presented target words. A GLMM was set up that was identical to the Fixation model except of the event type and word variable. For reasons of better interpretability, we omitted both variables and instead included a dummy coded variable that differentiated between the three target words and the event types *noRestructuring_AHA* and *Restructuring_AHA*.

Both equivalence tests were performed on the parameters of both models based on the equivalence bounds of .1 times the standard deviation of the respective outcome variable (see recommendation for .5 times the standard deviation for a point estimator in Norman, Sloan & Wyrwich, 2003).

MRI data acquisition, preprocessing and general linear analysis

Brain images were acquired during a 3 Tesla Siemens Magnetom Skyra MRI scanner system (Siemens Medical Systems, Erlangen, Germany) with a 20-channel radiofrequency head coil. Functional images were acquired using a T2*-weighted echo planar imaging (EPI) sequence sensitive to blood oxygen level dependent (BOLD) contrast (TR = 2400 ms, TE = 30 ms, image matrix = 64 × 64, voxel size = 3.0 mm × 3.0 mm × 3.0 mm, flip angle = 80°, FOV = 216 mm, 36 axial slices). Structural images were acquired during a three-dimensional T1-weighted magnetization prepared gradient-echo sequence (MPRAGE) (repetition time (TR) = 2500 ms; echo time (TE) = 2.12 ms; TI = 1100 ms, acquisition matrix = 256 × 256 × 192, flip angle = 9°; FOV = 240 mm, voxel size = 0.8 mm × 0.8 mm × 0.9 mm).

Image Processing. All fMRI data analyses were performed in SPM12 (Wellcome Department of Cognitive Neurology, London, UK). The fMRI images were slice-time corrected, realigned and coregistered to the individual structural image. Structural images were segmented into cerebral fluid, white and gray matter and normalized using the derived deformation field to the Montreal Neurological Institute (MNI) space. The deformation field was applied then also to the functional images that were subsequently smoothed with an 8-mm FWHM-Gaussian filter.

Statistical analyses. As implemented in SPM, all first and second level analyses were conducted in the framework of general linear models (GLMs). Two first level analyses were conducted (subsequently named first level analyses #1 and #2). The first one (#1) includes all trials (27 subjects). To exclude solution time as possible confound for differences in BOLD response between *no/Restructuring_AHA* events, a second first level model (#2) was set up with less trials (solution time did not differ anymore between *no/Restructuring_AHA* events, 22 subjects). See table 3 for list of all second level GLMs (appendix).

First level analysis #1: A total of 20 separate regressors were created for the event related first level analysis (the numbers in brackets [1-20] indicate the number of the respective regressors): We modeled the onset of each CRA problem [1] and all search button presses together as a separate onset regressor [2]. Furthermore, we

modeled the onset of the solution [3-6] for the four event types (*no/Restructuring_AHA*, *no/Restructuring_noAHA*). To get an estimate of the influence of solution time on the BOLD signal, we additionally included solution time as parametric modulator [7-10] for the four event types during solution. We further modeled all other remaining button presses (e.g. button presses regarding the 5 response options) using a separate nuisance onset regressor [11]. All eleven onset regressors were convolved with a canonical hemodynamic response function (HRF) and its first temporal derivative. Additionally, we separately modeled the mean for each of the three runs [12-14] and six motion parameters [15-20]. To remove baseline drifts, a high-pass filter with a 128 sec cutoff period was applied. Finally for each individual, simple contrast images were calculated from the beta weights of the onset regressors [1-11].

Second level analysis #1a & #1b: First, we included the con images corresponding to the four event types (*no/Restructuring_AHA*, *no/Restructuring_noAHA*) during solution in a random effects analysis using SPM's full factorial design (henceforth called #1a). The specific contrasts are reported in the Results section. Second, to estimate the effect of solution time on the results of especially the contrast *Restructuring_AHA>noRestructuring_AHA*, we conducted a separate SPM's full factorial design with only the betas of the parametric modulators of the four event types during solution (henceforth called #1b). The spatial overlap between the contrast *Restructuring_AHA>noRestructuring_AHA* (green) and the significant clusters of the parametric modulator solution time for *Restructuring_AHA* events (red) is depicted in figure 6 in the appendix.

To correct for multiple comparisons, a cluster threshold of $p < .001$ ($t > 3.19$, uncorrected) was applied at the voxel level. The resulting clusters were whole brain FWE corrected for multiple comparisons at $p < .05$.

First level analysis #2: Based on the results of the parametric modulator, we explored whether the observed differences in mean BOLD response from the contrast *Restructuring_AHA>noRestructuring_AHA* were only due to the longer solution time in the restructuring condition. Therefore, we conducted an additional first level analysis (#2) that was identical to #1 with two exceptions: We omitted the estimation of parametric modulators and included only a reduced number of trials for *no/Restructuring_AHA* events. That is to say, for each subject we excluded trials for *no/Restructuring_AHA* events until both event types would not significantly differ in solution time anymore ($p > .7$). Furthermore, we excluded subjects from the subsequent analysis if less than 5 trials per event type (*no/Restructuring_AHA*) would remain. This led to an exclusion of another 5 subjects (final subject size $n=22$) and solution time between both event types over all 22 subjects revealed no significant difference anymore ($t(629.3)=0.035$, $p=.403$). On average, 17.6 ($SD=4.60$) trials for the *noRestructuring_AHA* and 10.68 ($SD=4.23$) trials for the *Restructuring_AHA* condition remained per person. Note, for the event types *no/Restructuring_noAHA*, we used all available trials per subject because there were not enough trials to adjust

for differences in solution time and because these event types were not theoretically relevant given our hypotheses.

A total of 16 separate regressors were created (the numbers in brackets [1-16] indicate the number of the respective regressors): We modeled the onset of each CRA problem [1] and all search button presses as a separate onset regressor [2]. Furthermore, we modeled the onset of the solution [3-6] for the four event types [*no/Restructuring_AHA*, *no/Restructuring_noAHA*]. Note, for *no/Restructuring_AHA* events we used a reduced number of trials adjusted for solution time.

We further modeled all other remaining button presses (e.g. button presses regarding the 5 response options) using a separate nuisance onset regressor [7]. All seven onset regressors were convolved with a canonical hemodynamic response function (HRF) and its first temporal derivative. Additionally, we separately modeled the mean for each of the three runs [8-10] and six motion parameters [11-16]. A high-pass filter with a 128 sec cutoff period was applied to remove baseline drifts. Finally, simple contrast images were calculated from the beta weights of the onset regressors [1-6] for each individual.

Second level analysis #2a & #2b: For reasons of comparability, we used the same full factorial design as in #1 including the four event types (*noRestructuring_AHA*, *Restructuring_AHA*, *noRestructuring_noAHA*, *Restructuring_noAHA*) during solution (henceforth called #2a). We expected a substantial reduction of power due to a reduction in subject and trial number. To avoid type II errors, we adopted the more lenient but still widely accepted cluster threshold of $p < .005$ at the voxel level, ($t > 2.96$, uncorrected) and the resulting clusters were again whole brain FWE corrected for multiple comparisons at $p < .05$ (Liebermann & Cunningham, 2009).

Two clusters (left and right anterior insula/ IFG) from the analysis #1 for the contrast *Restructuring_AHA > noRestructuring_AHA* did not survive in analysis #2a. However, for sake of transparency given that less statistical power was expected, we decided to list the corresponding t-values. For confirmatory reasons, we additionally performed SPM's flexible factorial model which was identical to the full factorial #2a (four event types (*noRestructuring_AHA*, *Restructuring_AHA*, *noRestructuring_noAHA*, *Restructuring_noAHA* during solution) but included an additional random subject factor (henceforth called #2b). We reasoned that that the flexible factorial model in SPM is more sensitive to detect differences between both event types (*no/Restructuring_AHA*) as it corrects for violations of the IID assumption (of independent and identically distributed random variables) that were expected due to fewer trials and subjects. Here, we adopted the usual cluster inducing threshold of $p < .001$ ($t > 3.19$, uncorrected) at the voxel level. The resulting clusters were whole brain FWE corrected for multiple comparisons at $p < .05$.

All anatomical areas were determined using the AAL atlas (Tzourio-Mazoyer et al., 2002) based on the peak voxels from the random effects analysis. The reported coordinates correspond to the Montreal Neurological Institute (MNI) coordinate system (MNI) (Evans et al., 1993). Finally, for visualization of the significant clusters,

we superimposed the statistical map from the respective contrast on an averaged normalized structural image of all participants.

Results

Behavioral results

Frequency

The frequencies of solutions per event type (*no/Restructuring_AHA*, *no/Restructuring_noAHA*) for both samples are shown in figure 2.

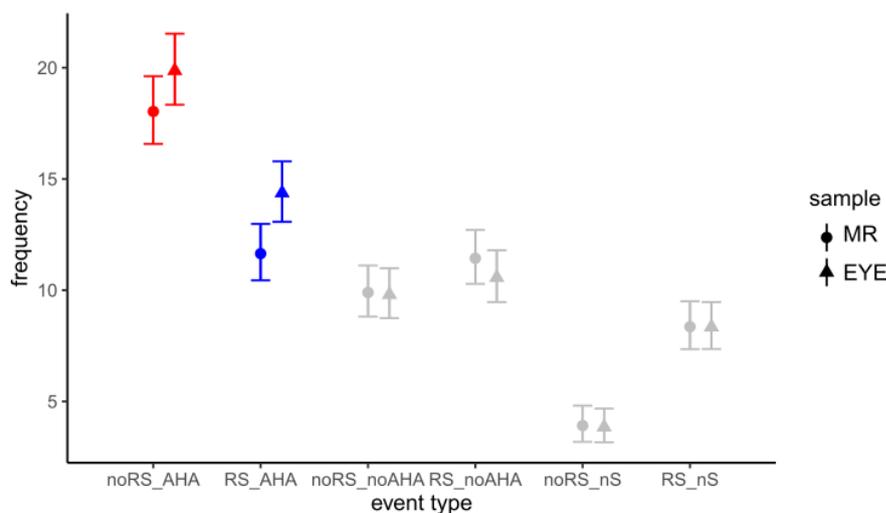


Figure 2. Frequency per event type (raw values). Frequency = average frequency of respective event type per subject. noRS_AHA = *noRestructuring_AHA* events; RS_AHA = *Restructuring_AHA* events; noRS_noAHA = *noRestructuring_noAHA* events; RS_noAHA = *Restructuring_noAHA* events; noRS_nS = CRA problems that were presented with a more closely related prime and not solved; RS_nS = CRA problems that were presented with a less closely related prime (restructuring condition) and not solved. Frequency per event type was depicted separately for the MR sample (round shape) and for the EYE sample (triangle shape). Error bars represent 95% confidence intervals. Solutions without AHA! experience and unsolved problems are depicted in gray because they were not part of the hypotheses and only plotted for sake of completeness.

Behavioral model

The influence of the event type on accuracy, solution time, search and search effort for both samples is shown in figure 3, their marginal means can be found in tables 4-5 and the respective model estimates can be found in tables 6-9 (appendix).

Solution time. Solution time was significantly shorter for *noRestructuring_AHA* compared to *Restructuring_AHA* events in the MR sample ($t(1442.61) = -8.31$, $p < .001$) and in the EYE sample ($t(1510) = -9.31$, $p < .001$). *NoRestructuring_AHA* events were also solved significantly faster than *noRestructuring_noAHA* in the MR sample ($t(1466.94) = -4.38$, $p < .001$) as well as in the EYE sample ($t(1558) = -11.62$, $p < .001$). Furthermore, solution time for

noRestructuring_AHA events was also significantly faster than for *Restructuring_noAHA* events in the MR sample ($t(1465.63)=-11.56, p<.001$) and in the EYE sample ($t(1532)=-20.65, p<.001$) (see fig. 3, upper left panel).

Accuracy. On average *noRestructuring_AHA* events were solved significantly more accurately compared to *Restructuring_AHA* events in the MR sample ($z=2.81, p<.05$) but not in the EYE sample ($p>.16$). Furthermore, accuracy was higher for *noRestructuring_AHA* events compared to *noRestructuring_noAHA* events in the MR sample ($z=4.51, p<.001$) as well as in the EYE sample ($z=4.36, p<.001$). Similarly, significantly more problems were solved correctly for *noRestructuring_AHA* compared to *Restructuring_noAHA* events in the MR sample ($z=9.11, p<.001$) and in the EYE sample ($z=8.58, p<.001$) events (see fig. 3, upper right panel).

Search. On average, participants pressed less often the search button for *noRestructuring_AHA* compared to *Restructuring_AHA* events in the MR sample ($z=-6.42, p<.001$) and in the EYE sample ($z=-3.60, p<.001$). Less search buttons were also pressed for *noRestructuring_AHA* compared to *noRestructuring_noAHA* events in the MR sample ($z=-9.59, p<.001$) and in the EYE sample ($z=-6.78, p<.001$). Similarly, less search buttons were pressed for *noRestructuring_AHA* compared to *Restructuring_noAHA* events in the MR sample ($z=-15.60, p<.001$) and in the EYE sample ($z=-10.86, p<.001$) (see fig. 3, lower left panel).

Search effort. Participants reported less effort to search for a solution for *noRestructuring_AHA* compared to *Restructuring_AHA* events in the MR sample ($t(1572.13)=-4.45, p<.001$) and in the EYE sample ($t(1523)=-4.47, p<.001$). Similarly, participants reported significantly less search effort for *noRestructuring_AHA* compared to *noRestructuring_noAHA* events in the MR sample ($t(1619.67)=-9.24, p<.001$) and in the EYE sample ($t(1567)=-9.36, p<.001$). Finally, *noRestructuring_AHA* events were also associated with less search effort compared to *Restructuring_noAHA* events in the MR sample ($t(1598.36)=-17.10, p<.001$) and in the EYE sample ($t(1546)=-17.04, p<.001$) (see, fig. 3, lower right panel).

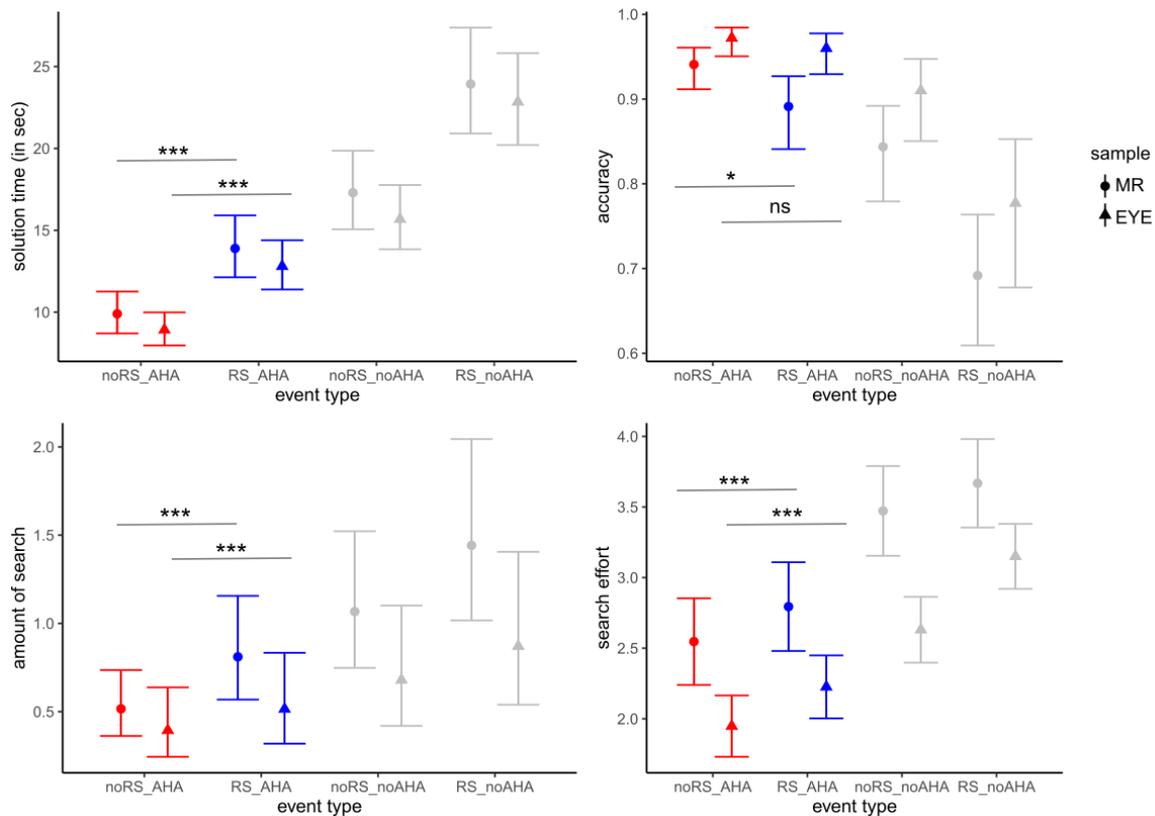


Figure 3. Behavioral models – Influence of event type on solution time, accuracy, amount of search and search effort demonstrated for both samples. Asterisks represent p-values: *** = $p < .001$, * = $p < .05$; ns = not significant; noRS_AHA = *noRestructuring_AHA* events; RS_AHA = *Restructuring_AHA* events; noRS_noAHA = *noRestructuring_noAHA* events; RS_noAHA = *Restructuring_noAHA* events; error bars represent 95% confidence intervals. Respective values per event type are model estimates and were depicted separately for the MR sample (round shape) and for the EYE sample (triangle shape). Solutions without AHA! experience are depicted in gray because they were not directly part of the hypotheses and only plotted for sake of completeness.

Eye-tracking results

Pupil response model

The pupil response as a function of the event type is shown in figure 4 and table 10 (appendix). The pupil response model with solution time as covariate did not reveal a statistically significantly better model fit than the pupil response model without solution time ($\chi^2(1)=3.09$, $p=.079$). Furthermore, none of the results from the pupil response model without solution time changed significantly due to inclusion of solution time as additional covariate. Therefore, all subsequent main effects and post-hoc analyses will be reported from the pupil response model without solution time.

There was a main effect of solution phase suggesting that pupil response was higher during solution than during trial start ($t(3076.94)=14.65$, $p<.001$). Furthermore, pupil dilation in general and irrespective of the solution phase was higher for *noRestructuring_AHA* events compared to *Restructuring_AHA* events ($t(3080.13)=-2.78$, $p<0.01$) but not to *noRestructuring_noAHA* ($p>0.07$) and *Restructuring_noAHA* ($p>0.21$) events. There was a significant solution phase and event type interaction

($t(3076.94)= 2.03, p<0.05$) for *Restructuring_AHA* vs. *noRestructuring_AHA* events. A visual inspection of figure 4 and a post-hoc analysis revealed that the pupil was only more strongly dilated for *noRestructuring_AHA* compared to *Restructuring_AHA* events during trial start ($t(1528.65)=2.75, p<0.01$) [as well as for *noRestructuring_noAHA* ($t(1547.27)=2.39, p<0.05$) and *Restructuring_noAHA* events ($t(1540.62)= 2.09, p<0.05$) during trial start]. However during solution, *Restructuring_AHA* and *noRestructuring_AHA* did not significantly differ ($p>.9$) anymore but pupil size was more strongly dilated for solutions with accompanied AHA! (*Restructuring_AHA, noRestructuring_AHA*) compared to solutions without AHA! (*noRestructuring_noAHA, Restructuring_noAHA*) ($t(1547.68)=3.01, p<.01$). The difference in pupil response (mean difference = $-.004, SE = 0.043; t(1527)=-.089, p=0.93$) between both solutions accompanied by an AHA! experience during solution was not significantly different from zero. According to the equivalence test, the 95% confidence interval [CI: $-.09; .08$] of the parameter values that distinguishes between both solutions with accompanied AHA! experience fell inside the region of practical equivalence [$-0.09; 0.09$].

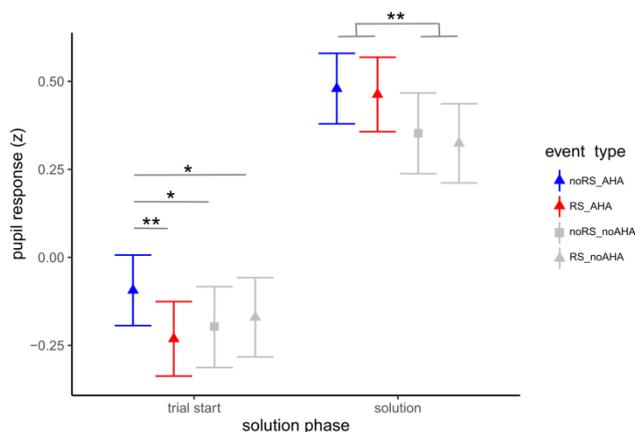


Figure 4. Pupil response model – Influence of event type and solution phase on pupil response. Asterisks represent p-values: ** = $p<.01$, * = $p<.05$; noRS_AHA = *noRestructuring_AHA*; RS_AHA = *Restructuring_AHA*, noRS_noAHA = *noRestructuring_noAHA*; RS_noAHA = *Restructuring_noAHA*; error bars represent 95% confidence intervals. Respective values per event type and solution phase are estimates from the pupil response model. Solutions without AHA! experience are depicted in gray because they were not directly part of the hypotheses and only plotted for sake of completeness.

Fixation model

The number of fixations and how they are distributed among the presented words (prime, first, second and third target word) as a function of the event type within the last two seconds before the solution are demonstrated in figure 5. The fixation model with solution time does not differ from the fixation model without solution time ($\chi^2(1)=0, p=1$). Therefore, all subsequent main effects and post-hoc analyses will be reported from the fixation model without solution time. The model estimates can be viewed in table 11 (appendix).

There is an overall main effect for the number of fixations on the first target word as it is more often fixated than the second ($z=3.84, p<.001$) and third ($z=4.52, p<.001$)

target word and the prime ($z=15.99$, $p<.001$). Furthermore, in *Restructuring_AHA* events more overall fixations were made compared to *noRestructuring_AHA* events ($z=2.93$, $p<.01$), while there was no significant evidence for more fixations during this time window for *Restructuring_AHA* compared to *noRestructuring_noAHA* ($p>.33$) and *Restructuring_noAHA* ($p>.09$) events. Significant interactions between the presented word and the event type can be observed in table 11 (appendix).

A posthoc analysis revealed, that there was no evidence of the first target word to be fixated more often for *Restructuring_AHA* events compared to all other events ($p>.19$). Importantly however, only in *Restructuring_AHA* events the first target word was relatively more often fixated when compared to the second ($t(6317)=3.84$, $p<0.05$) and third ($t(6317)=4.52$, $p<0.001$) target word. There was no evidence that the first target word was relatively more often fixated compared to the second and third target word in neither *noRestructuring_noAHA* ($p>.81$), *Restructuring_noAHA* ($p>.05$) nor *noRestructuring_AHA* ($p>.99$) events.

Finally, we did not find evidence that the amount of fixations for the first target word is statistically different to the other target words for *noRestructuring_AHA* events (first vs. second target word: mean difference = 0.015, $SE=.054$, $z=-.274$, $p=1$; first vs. third target word: mean difference = $-.044$, $SE=.054$, $z=-.809$, $p=1$). The equivalence test suggests that the 95% confidence interval of the mean difference between the first vs. second target word [95%CI: $-.12;0.09$] lies entirely inside the region of practical equivalence [$-.13, .13$]. However, only 76.24% of the 95% confidence interval regarding the difference between the first vs. third target word [95% CI: $-.03; .18$] lies inside the region of practical equivalence.

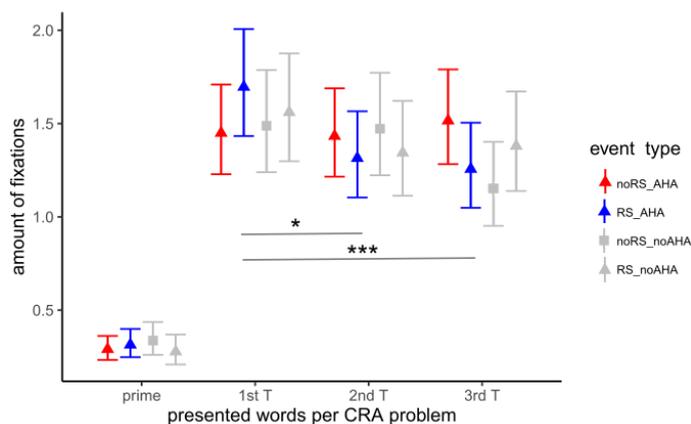


Figure 5. Fixation model – Influence of event type and presented words (1st, 2nd, 3rd target word or prime word) on the amount of fixations. Asterisks represent p-values: *** = $p<.001$, * = $p<.05$; 1st T = 1st target word, 2ndT = 2nd target word, 3rdT = 3rd target word. The amount of fixations refer to a time window of the last two seconds before the participants pressed the solution button. The first target word (*drop*) is the one that needs to be verbally restructured when presented with a less related prime (*backdrop*). noRS_AHA = *noRestructuring_AHA*; RS_AHA = *Restructuring_AHA*, noRS_noAHA = *noRestructuring_noAHA*; RS_noAHA = *Restructuring_noAHA*; error bars represent between subject 95% confidence intervals. Respective values per event type are estimates from fixation model. Solutions without AHA! experience are depicted in gray because they were not directly part of the hypotheses and only plotted for sake of completeness.

fMRI results

Second level model#1a - Main effects for event types during solution without adjusted solution time.

We tested whether *Restructuring_AHA* events involve more cognitive control during solution specifically in areas related to restructuring compared to *noRestructuring_AHA* events. The random effects analysis of the contrast *Restructuring_AHA>noRestructuring_AHA* during solution revealed a significant activation in the following three clusters: the left anterior insula extending into the left inferior frontal gyrus, the right anterior insula and left anterior cingulum extending into bilateral superior medial prefrontal gyrus (for details see table 2). The significant clusters from the latter contrast are shown in fig. 6.

The reverse contrast (*noRestructuring_AHA>Restructuring_AHA*) did not reveal any significant results.

Furthermore and for the sake of completeness, contrasting both solutions without accompanied AHA! experience did not lead to any significant results for the contrast *Restructuring_noAHA > noRestructuring_noAHA*. The reverse contrast *noRestructuring_noAHA > Restructuring_noAHA* revealed a significant cluster in the bilateral medial orbital frontal gyrus [peak voxel x;y;z(MNI) = 4,36,-12; $t=4.99$; cluster size: 127 voxels; $p<.01$].

Second level model #1b- Influence of parametric modulator solution time on BOLD response for Restructuring_AHA events

The behavioral results revealed that *Restructuring_AHA* problems take longer to solve than *noRestructuring_AHA* problems. Therefore, we first explored whether the observed clusters from the *Restructuring_AHA>noRestructuring_AHA* contrast spatially overlap with those voxels that significantly correlate with solution time for *Restructuring_AHA* events. Clusters of BOLD response were significantly correlated with solution time in the right inferior frontal gyrus, the right anterior insula, the right supplementary motor area, the left inferior frontal gyrus (pars orbitalis), the right superior frontal gyrus, bilateral caudate nucleus, the middle occipital gyrus and the left anterior cingulum. All clusters are detailed in table 2. The spatial overlap between the contrast *Restructuring_AHA>noRestructuring_AHA* and the significant clusters of the parametric modulator solution time for *Restructuring_AHA* events is depicted in figure 7, appendix.

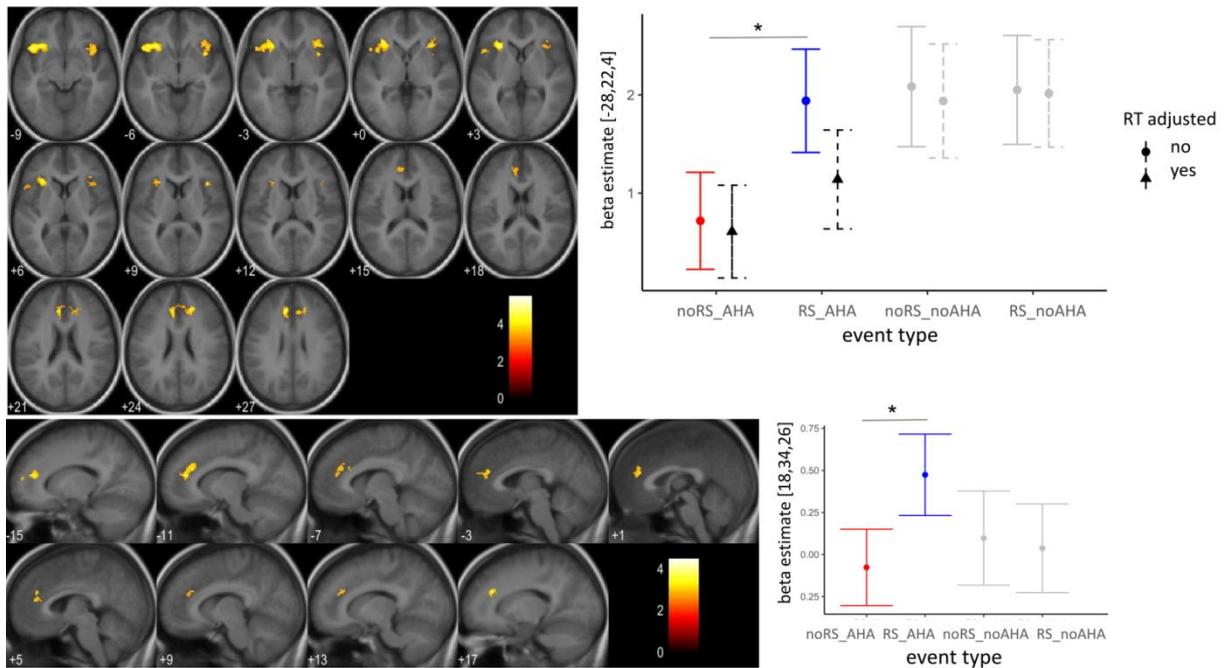


Figure 6. fMRI results for contrast Restructuring_AHA > noRestructuring_AHA.

Upper left panel: Significant BOLD response for contrast *Restructuring_AHA > noRestructuring_AHA* not adjusted for solution time between both event types (FullFactorial #1). RT adjusted = solution time adjusted: reduced number of onsets was included in the GLM so that solution time between *no/Restructuring_AHA* events is no longer significantly different ($p > .4$). Threshold for visualization: voxel-based $p < .001$; cluster-based $p < .05$, FWE corrected.

Upper right panel: Plot represents beta values for peak voxel in left anterior insula [-28, 22, 4] per event type for this contrast. Solutions without AHA! experience are depicted in gray because they were not part of the hypotheses and only plotted for sake of completeness.

Lower left panel: Significant BOLD response for contrast *Restructuring_AHA > noRestructuring_AHA* adjusted for solution time between both even types (second level model #2a). Threshold for visualization: voxel-based $p < .005$; cluster-based $p < .05$, FWE corrected).

Lower right panel: Plot represents beta values for peak voxel [-18, 34, 26] from ACC cluster per event type for this contrast.

Error bars in upper and lower right panel present 90% confidence intervals and asterisks represent a significant difference.

Second level model #2a - Main effects for contrast Restructuring_AHA > noRestructuring_AHA during solution with adjusted solution time

The cluster with the left anterior cingulum extending into left superior medial frontal gyrus and white matter was also significant with the reduced number of trials and subjects (see table 2, fig. 6, lower right panel). However, the two anterior insular/IFG clusters failed to reach significance when corrected for multiple comparisons. The reversed contrast did not reveal any significant clusters.

Second level model 2b# - Main effects for contrast Restructuring_AHA > noRestructuring_AHA during solution with adjusted solution time

The contrast *Restructuring_AHA > noRestructuring_AHA* revealed two significant clusters in the anterior cingulum. One cluster extended into superior medial frontal gyrus [peak voxel $x;y;z$ (MNI) = -12,40,18; $t=4.06$; cluster size: 145, voxels; $p < .001$] and the other one into white matter [peak voxel $x;y;z$ (MNI) = 2,44,6; $t=4.11$;

cluster size: 152 voxels; $p < .001$]. The reverse contrast did not reveal any significant clusters.

Table 2.

fMRI results from the second level models #1a, #1b and #2a during solution

	Side	Peak voxel MNI coordinates (mm)			Cluster size (voxel)	P-value of cluster size (FWE)	Peak t score
		x	y	z			
#1a - Restructuring_AHA>noRestructuring_AHA events							
ACC extending to SMFG	L/R	18	34	26	520	<.001	5.50 [4.46*]
AIC extending to IFG	L	-32	24	2	581	<.001	4.64 [2.96*]
AIC extending to IFG, pars orbitalis	R	36	10	-6	307	<.001	4.20 [2.21*]
#1b - brain areas correlating with solution time for Restructuring_AHA events							
ACC	L	-4	16	28	167	<.001	4.16
IFG	R	42	24	16	642	<.001	5.88
AIC	R	34	28	0	237	<.001	5.68
SMA	R	6	14	48	645	<.001	5.32
IFG, pars orbitalis	L	-30	30	-8	316	<.001	5.26
superior frontal gyrus	R	28	46	20	157	<.001	4.82
middle occipital gyrus	R	34	-74	32	311	<.001	4.35
caudate	L/R	14	10	12	180	<.001	4.21
#2a - Restructuring_AHA>noRestructuring_AHA events							
ACC extending to SMFG, and left middle frontal gyrus	L/R	18	34	26	614	<.005	4.46

Note. #1 – analysis with all trials included; #2 – analysis with less trials included but similar trial length between no/Restructuring_AHA events ($p < .4$); clusters are whole-brain FWE corrected for multiple comparisons at $p < .05$ and voxelwise threshold was $p < .001$ for #1 and $p < .005$ for #2 (see Liebermann & Cunningham, 2009); L = left; R = right; size (voxel) = cluster size, peak t score = maximum t value of significantly activated clusters; ACC = anterior cingulum; AIC = anterior insular cortex; SMFG = superior medial frontal gyrus; IFG = inferior frontal gyrus; SMA = supplementary motor area; squared brackets: t value for same voxel from second level model #2a (less included trials for no/Restructuring_AHA events to adjust for solution time); asterisk signifies that the values for the respective peak voxel are significant (uncorrected).

Discussion

In this article, we tested the hypothesis that solutions for verbal insight problems accompanied by an AHA! experience can also occur without prior restructuring. In other words, we argue that the link between the AHA! experience and restructuring is not as close as previously assumed. Moreover, we tested the derived hypothesis that dissociable processes underlie AHA! solutions with and without prior

restructuring. Results from two independent samples demonstrate that AHA! solutions do exist without prior restructuring (*noRestructuring_AHA* events) and that this event type is, in fact, the most frequent one (35-40% of all solved CRA problems).

As potential mechanism, we proposed that this phenomenon occurs when the problem is not immediately obvious but spreading activation starting from the target words activates related concepts that in turn activate the solution word. As a consequence, the solution word is activated strongly enough to become aware. In contrast, we proposed a different solution process for problems solved with an AHA! experience but *with* prior restructuring (*Restructuring_AHA* events). Spreading activation from the target words at first does not suffice to activate the solution because too many solution irrelevant associations are activated. Spreading activation can only activate the solution after involvement of restructuring or cognitive control. Cognitive control in this specific context means redirecting the attentional focus to those solution-irrelevant associations and subsequently inhibiting them as well as activating other (solutions-relevant) associations.

Derived from these two assumed mechanisms, we argued that the AHA! experience does not underlie a single neurocognitive process. If that was true, then both types of AHA! solutions (*noRestructuring_AHA* events) should exhibit dissociable behavioral, neuronal and eye-tracking related signatures. Taken together, our results support this hypothesis:

We could show converging evidence from both samples that *noRestructuring_AHA* events are solved faster, involve less search and search effort compared to *Restructuring_AHA* events (and solutions without AHA! experience). Concerning accuracy, more problems were solved correctly for *noRestructuring_AHA* compared to *Restructuring_AHA* events only in the MR but not in the EYE sample. This suggests that accuracy is not a stable predictor for differences in solutions with AHA! experience but only for differences between solutions with and without accompanied AHA! experience (see discussion further below). In sum, the behavioral results suggest that *Restructuring_AHA* events involve an additional source of task difficulty compared to *noRestructuring_AHA* events.

Furthermore, we provide evidence from fixation patterns two seconds before the solution that this additional source of task difficulty for *Restructuring_AHA* events may specifically refer to the need of reinterpreting the first target word and hence to restructuring. Only for *Restructuring_AHA* events, participants fixated the first target word significantly more often than the other target words. In contrast, for *noRestructuring_AHA* events the solution did not seem to depend on the first target word as it was fixated equally often as the second target word. Similarly, no significant difference could be observed in the amount of fixations between the first and the third target word for *noRestructuring_AHA* events (note, however, the amount of fixations for the first and the third target words was also not statistically equivalent).

In line with the behavioral and fixation pattern results, we provide evidence that *Restructuring_AHA* events during solution involve more prefrontal brain areas (ACC) associated with cognitive control compared to *noRestructuring_AHA* events. The ACC, as part of the salience network, is one of the brain areas that is most

consistently associated with insight solutions in general (see Sprugnoli et al., 2017; Dietrich & Kanso, 2010). Note, there are other regions like the right anterior temporal gyrus (Tik et al., 2018; Jung-Beeman et al., 2004) as well as the medial temporal lobe including the hippocampus and the amygdala that have also been associated with insight problem solving (Kizilirmak et al., 2016 & 2019; Becker et al., 2020b; Ludmer et al., 2011). However, the ACC has additionally been identified in the context of restructuring and interpreted as mediator of breaking one's mental set to form novel task-related associations (Zhao et al., 2013; Luo et al. 2004b, Wu et al., 2013). That is to say, the ACC may function as a "warning system" in the sense that activated solution-irrelevant associations are hindering the retrieval of the correct solution. This interpretation is in line with the ACC's associated function of conflict monitoring during cognitive control (Botwinick et al., 2004; Kerns et al., 2004). Apart from the ACC, we also found increased IFG and anterior insular activation for the contrast *Restructuring_AHA > noRestructuring_AHA*. The IFG has been associated with inhibiting those solution-irrelevant problem elements and activating other (among them solution-relevant) elements necessary to retrieve the correct solution from memory (Becker et al., 2020a; Wu et al., 2013; Tang et al., 2015). However, BOLD response in these areas was correlated with solution time (see fig. 7, appendix). When correcting for the difference in solution time, activity in the IFG (and anterior insula) related to the *Restructuring_AHA* condition decreased as illustrated in figure 6 (upper left panel) and the difference did not survive the correction for multiple comparisons anymore. This may be due to the loss in power given that fewer trials and subjects were included in the solution time adjusted GLM. Hence, further research is necessary to disentangle both effects in IFG and to investigate to which degree increased solution time in *Restructuring_AHA* events reflects restructuring relevant and irrelevant processes.

In sum, the fMRI evidence supports the assumption that *Restructuring_AHA* events involve more cognitive control during solution than *noRestructuring_AHA* events as reflected in increased BOLD response in ACC. This increased response may reflect a "breaking of the mental set" as in an error signal towards the solution-irrelevant meaning of the first target word.

At last, there is first evidence from pupil response that *noRestructuring_AHA* events may be processed earlier compared to *Restructuring_AHA* events (and solutions without AHA! experience) as measured by increased pupil dilation during trial start. Differences in pupil response have been related to recognition memory and lexical retrieval: Otero and colleagues found the pupil to be more strongly dilated when participants viewed old versus new items during recognition memory tests suggesting that the pupil dilation may reflect the strength of the underlying memory trace (Otero, Weekes & Hutton, 2011). In addition, Kafkas & Montaldi (2012) report larger pupil dilations related to recognition memory of visual object stimuli. Finally, Renner and Wlodyczak report greater pupil dilation shortly after stimulus onset related to lexical retrieval (Renner & Wlodyczak, 2017).

In sum, the presented results together suggest that 1) there can be solutions with AHA! experience but *without* prior restructuring; 2) solutions accompanied by an AHA! experience exhibit dissociable neural, behavioral and eye-tracking related processes as a function of restructuring and therefore are supported by different neurocognitive processes.

The main question that arises from these results is the following. If an AHA! experience is not necessarily caused by restructuring, then what does it reflect?

It may just suffice that the solution is *perceived* as sudden and correct to experience an AHA! while the AHA! experience itself may be indifferent to the cause of why the problem is solved suddenly.

There is various evidence that the perceived correctness and suddenness or surprise as a related concept (Bilalić et al., 2019) are key to differentiate between solutions with and without AHA! experience: Accuracy has been consistently shown to be higher for solutions with compared to without an AHA! experience (Danek et al., 2014; Salvi et al., 2016; Becker et al., 2020b, Webb et al., 2016). Furthermore, pupil response has been associated with surprise (Preuschoff et al., 2011; Kloosterman et al., 2015). The results presented here demonstrate that during solution, pupil response is statistically equivalent between solutions with AHA! experience irrespective of the restructuring condition and pupil response is significantly higher than for solutions without AHA! experience. Hence, both solutions accompanied by an AHA! may have been perceived as similarly surprising and respectively more surprising than solutions without AHA! experience. A similar result has also been reported before by Salvi and colleagues (2019). Using a non modified CRA paradigm, they also report increased pupil response for solutions with compared to without AHA! experience during solution. Moreover, pupil response has also been associated with emotional arousal (Bradley et al., 2008; Partala & Surakka, 2003). Hence, the increased pupil response for both AHA! solutions (*no/Restructuring_AHA* events) could equally well reflect the positive affective response to the sudden occurrence of the solution.

However, if the AHA! experience only depends on the non-obvious solution to be *perceived* as sudden and correct irrespective of prior restructuring, why has this event type not been reported before? According to the representational change theory and its successors, it is assumed that a correct initial problem representation should progress smoothly and continuously to a solution not leading to an AHA! experience (see Danek, 2018; Ohlsson, 1992). Hence, a solution with accompanied AHA! experience to a problem whose initial problem representation was initially correct (and therefore needed no prior restructuring as in *noRestructuring_AHA* events) should not exist according to this theory. However, this very assumption does not necessarily hold in CRAs. Here, the solution is not immediately obvious² even if

² Note, there are CRA problems that are solved in two or less seconds. Most likely the solution word was directly activated by the target words. These kind of solutions are not considered as neither insight nor trial and error solutions.

the meaning of all target words would be presented in a solution-relevant way. The reason for this is because the solution word still needs to be retrieved from memory. We speculate that this additional retrieval process may take enough time (e.g. more than 2 seconds) for the solution to seem non-obvious at first and therefore produce a sense of surprise upon retrieval of the solution. This should especially be the case if the solution word is automatically activated via other concepts being activated first which then in turn activate the solution. In contrast in classical insight tasks, the solution is most likely not immediately obvious to the solver *because* the solver acquired an incorrect problem representation (requiring restructuring and subsequently leading to an AHA! experience). In sum, while in classical insight tasks an incorrect problem representation, restructuring and the AHA! experience coincide most of the time, this is not the case for CRAs. In CRAs, a solution can seem non-obvious at the beginning without initial incorrect problem representation due to a memory retrieval component.

Importantly, while the AHA! experience itself may be indifferent to the cause of why a solution was generated suddenly, the quality or subjective experience of the AHA! may be dependent on whether or not restructuring preceded the solution accompanied by AHA!.

The AHA! experience is not a unitary construct as behavioral, neural and eye-tracking related results from this study demonstrate. Furthermore, Danek and colleagues showed that the different dimensions of the AHA! experience like pleasure, suddenness and certainty can vary in magnitude and quality (Danek & Wiley, 2017). Hence, it is possible that an AHA! solution may be experienced with more pleasure and a higher degree of suddenness when the problem was restructured compared to when it was not. For this study, the AHA! experience was measured in a binary mode (absence or presence of AHA! experience) as it is also done with classical CRAs (Jung-Beeman et al., 2004, Salvi et al., 2016; Bowden & Jung-Beeman, 2007). Furthermore, when defining the concept of the AHA! experience for the participant (see Method's section), we stressed the cognitive component of the AHA! experience (suddenness, certainty) while not explicitly emphasizing on its emotional component (surprise, positive affect). With the available data, the above-mentioned different nuances in the AHA! experience including its emotional component cannot be differentiated as a function of prior restructuring. Therefore, future insight research using (modified) CRAs should assess the AHA! experience in a more nuanced way including its emotional component, the magnitude and quality of this subjective experience.

Of course, the question would still be *why* the solution with an accompanied AHA! is perceived as sudden and obviously correct. In this context, we proposed a spreading activation account that at least in the case of CRAs could explain why a solution suddenly becomes aware to the solver. Note, this account is closely related to earlier theories of unconscious processing in insight problem solving (Bowden & Beeman, 1998; Bowden, 1997). These proposed distinct CRA solution processes (both

including spreading activation as a key concept) are in line with the results. However, those processes are only descriptive. To falsify their implicit assumptions, a goal for future research should be to formalize those solution processes into a neural network model, simulate data and compare the results to the empirical data. First promising modelling approaches on creative problem solving can be found in the literature (Sun & Helie, 2010; Olteteanu et al., 2015; Kajic et al., 2017). However, this line of research is still vastly unexplored and just beginning to emerge. For example, we are not aware of any attempts to formalize or model restructuring.

One might argue that solutions with accompanied AHA! experience but without restructuring cannot be classified as insight anymore because the restructuring aspect is missing. Whether *noRestructuring_AHA* events can be classified as insight, depends on its definition. Researchers have distinguished insight from non-insight solutions with (e.g. Danek et al., 2020; Ohlsson, 1992) but also without (e.g. Bowden & Beeman, 1997) restructuring as necessary prerequisite. However, irrespective of which insight definition one adopts, the results of this study have important implications for further insight research especially when investigating CRAs. First, it cannot not be implied anymore that restructuring has taken place by just measuring the AHA! experience. Second, if a definition of insight including both restructuring and the AHA! experience is accepted, then additional effort should be made to assess both aspects of insight. But even if one were to accept a definition of insight without the concept of restructuring, a separate investigation of AHA! solutions with and without restructuring is advisable given their dissociable underlying neurocognitive processes.

References

- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, 22(3), 261-295.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390 – 412.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0 – 6.
- Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4), 602-607.
- Becker, M., Wiedemann, G., & Kühn, S. (2018). Quantifying insightful problem solving: A modified compound remote associates paradigm using lexical priming to parametrically modulate different sources of task difficulty. *Psychological Research*. <https://doi.org/10.1007/s00426-018-1042-3>.
- Becker, M., Sommer, T., & Kühn, S. (2020a). Inferior frontal gyrus involvement during search and solution in verbal creative problem solving: A parametric fMRI study. *NeuroImage*, 206, 116294.
- Becker, M., Sommer, T., & Kühn, S. (2020b). Verbal insight revisited: fMRI evidence for early processing in bilateral insulae for solutions with AHA! experience shortly after trial onset. *Human brain mapping*, 41(1), 30-45.
- Biemann, C., Heyer, G., Quasthoff, U., Richter, M., (2007). The Leipzig Corpora collection: monolingual corpora of standard size. In: *Proceedings of Corpus Linguistic 2007*. (Birmingham).
- Bilalić, M., Graf, M., Vaci, N., & Danek, A. H. (2019). When the Solution Is on the Doorstep: Better Solving Performance, but Diminished Aha! Experience for Chess Experts on the Mutilated Checkerboard Problem. *Cognitive science*, 43(8), e12771.
- Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Research*, 50(23), 2577 – 2587.
- Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in cognitive sciences*, 8(12), 539-546.
- Bowden, E.M. (1997). The effect of reportable and unreportable hints on anagram solution and the aha! experience. *Consciousness and Cognition*, 6, 545–573.
- Bowden, E. M., & Beeman, M. J. (1998). Getting the right idea: Semantic activation in the right hemisphere may help solve insight problems. *Psychological science*, 9(6), 435-440.
- Bowden, E.M., Jung-Beeman, M., Fleck, J., Kounios, J., (2005). New approaches to demystifying insight. *Trends Cogn. Sci.* 9 (7), 322–328.
- Bowden, E.M., Jung-Beeman, M., (2003). Normative data for 144 compound remote associate problems. *Behav. Res. Methods Instrum. Comput.* 35 (4), 634–639.
- Bowden, E.M., Jung-Beeman, M. (2007). Methods for investigating the neural components of insight. *Methods* 42 (1), 87–99.

- Bowers, K. S., Regehr, G., Balthazard, C., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive psychology*, 22(1), 72-110.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433 – 436.
- Danek, A. H., Williams, J., & Wiley, J. (2020). Closing the gap: connecting sudden representational change to the subjective Aha! experience in insightful problem solving. *Psychological research*, 84(1), 111-119.
- Danek, A. H., & Wiley, J. (2017). What about false insights? Deconstructing the Aha! experience along its multiple dimensions for correct and incorrect solutions separately. *Frontiers in psychology*, 7, 2077.
- Danek, A. H., Wiley, J., & Öllinger, M. (2016). Solving classical insight problems without aha! experience: 9 dot, 8 coin, and matchstick arithmetic problems. *The Journal of Problem Solving*, 9(1), 4.
- Danek, A. H. (2018). Magic tricks, sudden restructuring, and the Aha! experience: A new model of nonmonotonic problem solving. In *Insight* (pp. 51-78). Routledge.
- Danek, A. H., Fraps, T., von Müller, A., Grothe, B., & Öllinger, M. (2014). Working wonders? Investigating insight with magic tricks. *Cognition*, 130(2), 174–185.
- Dietrich, A., & Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological bulletin*, 136(5), 822.
- Duncker, K., & Lees, L. S. (1945). On problem-solving. *Psychological monographs*, 68(5).
- Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, 113(28), 7900 – 7905.
- Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., & Peters, T. M. (1993). 3D statistical neuroanatomical models from 305 MRI volumes. Nuclear Science Symposium and Medical Imaging Conference, 1993 IEEE Conference Record (pp. 1813 – 1817). IEEE.
- Gardner, W., Mulvey, E.P., Shaw, E.C., (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychol. Bull.* 118 (3), 392–404.
- Goel, V., (2014). Creative brains: designing in the real world. *Front. Hum. Neurosci.* 8 (241), 1–14.
- Hélie, S., & Sun, R. (2010). Incubation, insight, and creative problem solving: a unified theory and a connectionist model. *Psychological review*, 117(3), 994.
- Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., ... & Kounios, J. (2004). Neural activity when people solve verbal problems with insight. *PLoS biology*, 2(4).
- Kajić, I., Gosmann, J., Stewart, T. C., Wennekers, T., & Eliasmith, C. (2017). A spiking neuron model of word associations for the remote associates test. *Frontiers in psychology*, 8, 99.

- Kafkas, A., & Montaldi, D. (2012). Familiarity and recollection produce distinct eye movement, pupil and medial temporal lobe responses when memory strength is matched. *Neuropsychologia*, *50*(13), 3080-3093.
- Kaplan, C. A. & Simon, H. A., (1990). In search of insight. *Cognitive Psychology*, *22*(3), 374–419. [http://dx.doi.org/10.1016/0010-0285\(90\)90008-R](http://dx.doi.org/10.1016/0010-0285(90)90008-R).
- Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*(5660), 1023-1026.
- Kizilirmak, J. M., Thuerich, H., Folta-Schoofs, K., Schott, B. H., & Richardson-Klavehn, A. (2016). Neural correlates of learning from induced insight: A case for reward-based episodic encoding. *Frontiers in Psychology*, *7*:1693.
- Kizilirmak, J. M., Schott, B. H., Thuerich, H., Sweeney-Reed, C. M., Richter, A., Folta-Schoofs, K., & Richardson-Klavehn, A. (2019). Learning of novel semantic relationships via sudden comprehension is associated with a hippocampus-independent network. *Consciousness and Cognition*, *69*, 113–132.
- Knapen, T., de Gee, J. W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., & Theeuwes, J. (2016). Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PloS one*, *11*(5).
- Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, memory, and cognition*, *25*(6), 1534.
- Kloosterman, N. A., Meindertsma, T., van Loon, A. M., Lamme, V. A., Bonnef, Y. S., & Donner, T. H. (2015). Pupil size tracks perceptual content and surprise. *European Journal of Neuroscience*, *41*(8), 1068-1078.
- Kounios, J., & Beeman, M. (2014). The cognitive neuroscience of insight. *Annual review of psychology*, *65*.
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259-269.
- Levy, O., & Goldberg, Y. (2014a). Dependency-based word embeddings. In Proceedings of the 52nd annual meeting of the association for computational linguistics (pp. 302–308). Baltimore: ACL.
- Levy, O., & Goldberg, Y. (2014b). Linguistic regularities in sparse and explicit word representations. In Proceedings of the eighteenth conference on computational natural language learning (pp. 171–180). Baltimore: ACL.
- Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social cognitive and affective neuroscience*, *4*(4), 423-428.
- Ludmer, R., Dudai, Y., & Rubin, N. (2011). Uncovering camouflage: amygdala activation predicts long-term memory of induced perceptual insight. *Neuron*, *69*(5), 1002-1014.
- Luo, J., Niki, K., & Phillips, S. (2004b). The function of the anterior cingulate cortex (ACC) in the insightful solving of puzzles: The ACC is activated less when the structure of the puzzle is known. *Journal of Psychology in Chinese Societies*, *5*(2), 195-213.

- Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). *bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework*. *Journal of Open Source Software*, 4(40), 1541. [10.21105/joss.01541](https://doi.org/10.21105/joss.01541)
- Millar, R. B. (2011). *Maximum likelihood estimation and inference: with examples in R, SAS and ADMB* (Vol. 111). John Wiley & Sons.
- Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Medical care*, 582-592.
- Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. Keane & K. J. Gilhooly (Eds.), *Advances in the psychology of thinking* (pp. 1–44). London: Harvester-Wheatsheaf.
- Ohlsson, S. (1984). Restructuring revisited: II. An information processing theory of restructuring and insight. *Scandinavian journal of psychology*, 25(2), 117-129.
- Öllinger, M., Jones, G., Faber, A. H., & Knoblich, G. (2013). Cognitive mechanisms of insight: the role of heuristics and representational change in solving the eight-coin problem. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 931.
- Öllinger, M., Jones, G., & Knoblich, G. (2014). The dynamics of search, impasse, and representational change provide a coherent explanation of difficulty in the nine-dot problem. *Psychological research*, 78(2), 266-275.
- Olteteanu, A. M., & Falomir, Z. (2015). comRAT-C: a computational compound Remote Associates Test solver based on language data and its comparison to human performance. *Pattern Recognition Letters*, 67, 81-90.
- Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology*, 48(10), 1346-1353.
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12), 976-987.
- Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International journal of human-computer studies*, 59(1-2), 185-198.
- Preuschoff, K., Hart, B. M., & Einhauser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in neuroscience*, 5, 115.
- R Core Team. (2014). *R: A language and environment for statistical computing* (p. 2014). Vienna, Austria: R Foundation for Statistical Computing.
- Renner, L. F., & Włodarczak, M. (2017). When a Dog is a Cat and How it Changes Your Pupil Size: Pupil Dilation in Response to Information Mismatch. In *INTERSPEECH* (pp. 674-678).
- Salvi, C., Bricolo, E., Kounios, J., Bowden, E., & Beeman, M. (2016). Insight solutions are correct more often than analytic solutions. *Thinking & reasoning*, 22(4), 443-460.
- Salvi, C., Simoncini, C., Beeman, M. & Grafman, J. (2019, April). *Pupil-size and Microsaccades predict different problem-solving styles*. Poster session presented at the Conference for the Society for the Neuroscience of Creativity, San Francisco, CA.

- Schunn, C.D., & Dunbar, K. (1996). Priming, analogy, and awareness in complex reasoning. *Memory & Cognition*, 24, 271–284.
- Subramaniam, K., Kounios, J., Parrish, T. B., & Jung-Beeman, M. (2009). A brain mechanism for facilitation of insight by positive affect. *Journal of cognitive neuroscience*, 21(3), 415–432.
- Sprugnoli, G., Rossi, S., Emmendorfer, A., Rossi, A., Liew, S.-L., Tatti, E., di Lorenzo, G., Pascual-Leone, A., & Santarnecchi, E. (2017). Neural correlates of Eureka moment. *Intelligence*, 62, 99–118.
- SR RESEARCH.(2007). EyeLink user manual.Version 1.3.0 [Computer software manual]. Ottawa, ON: Author.
- Sternberg, R. J., & Davidson, J. E. (1995). *The nature of insight*. Cambridge: MIT Press.
- Tang, X., Pang, J., Nie, Q. Y., Conci, M., Luo, J., & Luo, J. (2015). Probing the cognitive mechanism of mental representational change during chunk decomposition: a parametric fMRI study. *Cerebral Cortex*, 26(7), 2991–2999.
- Thevenot, C., & Oakhill, J. (2008). A generalization of the representational change theory from insight to non-insight problems: The case of arithmetic word problems. *Acta Psychologica*, 129, 315 – 324.
- Tik, M., Sladky, R., Di BernardiLuft, C., Willinger, D., Hoffmann, A., Banissy, M. J., Bhattacharya, J., & Windischberger, C. (2018). Ultra-high-field fMRI insights on insight: Neural correlates of the Aha!-moment. *Human Brain Mapping*, 39(8), 3241–3252.
- Topolinski, S., & Reber, R. (2010). Gaining insight into the “Aha” experience. *Current Directions in Psychological Science*, 19(6), 402–405.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1), 273 – 289.
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, 14(5), 779–804.
- Webb, M. E., Little, D. R., & Cropper, S. J. (2016). Insight is not in the problem: Investigating insight in problem solving across task types. *Frontiers in psychology*, 7, 1424.
- Wu, L., Knoblich, G., & Luo, J. (2013). The role of chunk tightness and chunk familiarity in problem solving: evidence from ERPs and fMRI. *Human brain mapping*, 34(5), 1173–1186.
- Zhao, Q., Zhou, Z., Xu, H., Chen, S., Xu, F., Fan, W., & Han, L. (2013). Dynamic neural network of insight: a functional magnetic resonance imaging study on solving Chinese ‘chengyu’ riddles. *PloS one*, 8(3), e59351.

Appendix

Table 3.

List of GLMs for fMRI analysis

Model name	Onset regressors	sample size	adjusted for solution time?	voxel-based threshold
#1a	4 onset regressors (4 event types)	n=27	no	p<.001
#1b	4 PM solution time * 4 event types	n=27		p<.001
#2a	4 onset regressors (4 event types)	n=22	yes	p<.005
#2b	4 onset regressors (4 event types)	n=22	yes	p<.001

Note. The four event types refer to *no/Restructuring_AHA* and *no/Restructuring_noAHA* events during solution. PM = parametric modulator.

Table 4.

Marginal means of solution time and accuracy per event type

	Solution time (in sec)		Accuracy (in % correct)	
	MR sample	EYE sample	MR sample	EYE sample
noRS_AHA	9.9 (8.7 - 11.3)	9.0 (8.0 - 10.1)	94 (91 - 96)	97 (95 - 99)
RS_AHA	13.9 (12.1 - 15.9)	13.1 (11.6 - 14.7)	89 (84 - 93)	96 (93 - 98)
noRS_noAHA	17.3 (5.1 - 19.9)	15.6 (13.7 - 17.7)	84 (78 - 89)	91 (85 - 95)
RS_noAHA	23.9 (20.9 - 27.4)	22.8 (20.1 - 25.8)	69 (61 - 76)	78 (67 - 85)

Note. noRS_AHA = noRestructuring_AHA; RS_AHA = Restructuring_AHA; noRS_noAHA = noRestructuring_AHA; RS_noAHA = Restructuring_noAHA; the values represent marginal means and the values in parentheses are the 95% confident intervals.

Table 5.

Marginal means of search and search effort per event type

	Search		Search effort	
	MR sample	EYE sample	MR sample	EYE sample
noRS_AHA	0.39 (0.36 - 0.74)	0.43 (0.27 - 0.68)	1.95 (1.73 - 2.17)	1.98 (1.77 - 2.18)
RS_AHA	0.81 (0.57 - 1.16)	0.56 (0.35 - 0.90)	2.23 (2.00 - 2.45)	2.27 (2.05 - 2.48)
noRS_noAHA	1.07 (0.75 - 1.52)	0.73 (0.46 - 1.17)	2.63 (2.40 - 2.86)	2.69 (2.47 - 2.91)
RS_noAHA	1.44 (1.02 - 2.04)	0.94 (0.59 - 1.49)	3.15 (2.92 - 3.38)	3.21 (2.99 - 3.43)

Note. noRS_AHA = noRestructuring_AHA; RS_AHA = Restructuring_AHA; noRS_noAHA = noRestructuring_AHA; RS_noAHA = Restructuring_noAHA; search = number of search button presses per trial; the values represent marginal means and the values in parentheses are the 95% confident intervals.

Table6.

Behavioral model - Influence of event type on solution time (log)

log(solution time) <i>Predictors</i>	MR sample			EYE sample		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	2.63	2.52 – 2.75	<0.001	2.57	2.45 – 2.65	<0.001
event type: RS_noAHA	0.54	0.47 – 0.62	<0.001	0.56	0.46 – 0.65	<0.001
event type: noRS_AHA	-0.34	-0.41 – -0.27	<0.001	-0.37	-0.45 – -0.30	<0.001
event type: noRS_noAHA	0.22	0.14 – 0.30	<0.001	0.18	0.08 – 0.28	<0.001
Random Effects						
Tau	0.02 _{item}			0.04 _{item}		
	0.10 _{subj}			0.07 _{subj}		
Marginal R ² / Conditional R ²	0.206 / 0.419			0.197 / 0.367		

Note. Tau = random effect variance; CI = 95% confidence intervall. RS_noAHA = Restructuring_noAHA; noRS_AHA = noRestructuring_AHA, noRS_noAHA = noRestructuring_noAHA.

Table7.

Behavioral model - Influence of event type on accuracy

accuracy <i>Predictors</i>	MR sample			EYE sample		
	<i>OR</i>	<i>CI</i>	<i>p</i>	<i>OR</i>	<i>CI</i>	<i>p</i>
(Intercept)	8.19	5.28 – 12.69	<0.001	24.68	13.29 – 45.84	<0.001
event type: RS_noAHA	0.27	0.18 – 0.42	<0.001	0.14	0.08 – 0.24	<0.001
event type: noRS_AHA	1.94	1.22 – 3.08	0.005	1.47	0.85 – 2.54	0.170
event type: noRS_noAHA	0.66	0.41 – 1.06	0.085	0.40	0.22 – 0.72	0.002
Random Effects						
Tau	0.29 _{item}			0.59 _{item}		
	0.41 _{subj}			1.00 _{subj}		
Marginal R ² / Conditional R ²	0.123 / 0.275			0.139 / 0.420		

Note. OR = odds ratios. Tau = random effect variance; CI = 95% confidence intervall. RS_noAHA = Restructuring_noAHA; noRS_AHA = noRestructuring_AHA, noRS_noAHA = noRestructuring_noAHA.

Table8.

Behavioral model - Influence of event type on amount of search

Search <i>Predictors</i>	MR sample			EYE sample		
	<i>IRR</i>	<i>CI</i>	<i>p</i>	<i>IRR</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.81	0.57 – 1.16	0.249	0.56	0.35 – 0.90	0.015
event type: RS_noAHA	1.77	1.55 – 2.02	<0.001	1.67	1.45 – 1.94	<0.001
event type: noRS_AHA	0.64	0.55 – 0.73	<0.001	0.76	0.66 – 0.88	<0.001
event type: noRS_noAHA	1.30	1.13 – 1.51	<0.001	1.30	1.11 – 1.53	0.001
Random Effects						
Tau	0.01 _{item}			0.01 _{item}		
	0.88 _{subj}			1.49 _{subj}		
Marginal R ² / Conditional R ²	0.090 / 0.582			0.035 / 0.621		

Note. *IRR* = incidence rate ratio. Tau = random effect variance; CI = confidence intervall. RS_noAHA = Restructuring_noAHA; noRS_AHA = noRestructuring_AHA, noRS_noAHA = noRestructuring_noAHA.

Table 9.
Behavioral model - Influence of event type on search effort

Searcheffort <i>Predictors</i>	MR sample			EYE sample		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	2.79	2.53 – 3.06	<0.001	2.27	2.05 – 2.48	<0.001
event type: RS_noAHA	0.87	0.74 – 1.00	<0.001	0.95	0.79 – 1.10	<0.001
event type: noRS_AHA	-0.25	-0.36 – -0.14	<0.001	-0.29	-0.41 – -0.16	<0.001
event type: noRS_noAHA	0.68	0.54 – 0.82	<0.001	0.42	0.27 – 0.58	<0.001
Random Effects						
Tau	0.02 _{item}			0.04 _{item}		
	0.67 _{subj}			0.25 _{subj}		
Marginal R ² / Conditional R ²	0.124 / 0.499			0.144 / 0.340		

Note. Tau = random effect variance; CI = 95% confidence intervall. RS_noAHA = Restructuring_noAHA; noRS_AHA = noRestructuring_AHA, noRS_noAHA = noRestructuring_noAHA.

Table 10.
Pupil response model - Influence of event type and solution phase on pupil response

<i>Predictors</i>	Pupilresponse		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	-0.16	-0.26 – -0.06	0.002
event type: Restructuring_AHA	-0.12	-0.21 – -0.04	0.005
event type: noRestructuring_noAHA	-0.09	-0.19 – 0.01	0.076
event type: Restructuring_noAHA	-0.06	-0.15 – 0.04	0.220
time: solution	0.59	0.51 – 0.67	<0.001
Pupilsizebaseline	0.19	0.16 – 0.22	<0.001
event type:Restructuring_AHA:time: solution	0.13	0.00 – 0.25	0.042
event type:noRestructuring_noAHA:time: solution	-0.04	-0.18 – 0.09	0.551
event type:Restructuring_noAHA:time: solution	-0.09	-0.23 – 0.04	0.166
Marginal R ² / Conditional R ²	0.191 / 0.278		

Note. CI = 95% confidence intervall, p = P-value. The solution phase is divided into two phases: first three seconds after trial start and last second before solution.

Table 11.

Fixation model – Influence of event type and presented word on number of fixations per presented word within two seconds before solution

<i>Predictors</i>	Numberoffixations		
	<i>IRR</i>	<i>CI</i>	<i>p</i>
word: 2T	0.78	0.69 – 0.89	< 0.001
word: 3T	0.74	0.65 – 0.84	< 0.001
word: prime	0.18	0.14 – 0.22	< 0.001
event type: noRestructuring_AHA	0.84	0.75 – 0.94	0.003
event type: noRestructuring_noAHA	0.94	0.82 – 1.07	0.333
event type: Restructuring_noAHA	0.89	0.78 – 1.02	0.092
wordlength	1.04	1.02 – 1.06	< 0.001
word: 2T - event type: noRestructuring_AHA	1.29	1.09 – 1.52	0.003
word: 3T - event type: noRestructuring_AHA	1.45	1.22 – 1.72	< 0.001
word: prime - event type: noRestructuring_AHA	1.13	0.85 – 1.50	0.395
word: 2T - event type: noRestructuring_noAHA	1.10	0.90 – 1.33	0.367
word: 3T - event type: noRestructuring_noAHA	1.16	0.94 – 1.42	0.160
word: prime - event type: noRestructuring_noAHA	0.99	0.70 – 1.39	0.935
word: 2T - event type: Restructuring_noAHA	1.23	1.01 – 1.49	0.037
word: 3T - event type: Restructuring_noAHA	1.03	0.84 – 1.26	0.799
word: prime - event type: Restructuring_noAHA	1.30	0.95 – 1.78	0.107
Marginal R ² / Conditional R ²	0.453 / 0.517		

Note. word length: amount of letters per most fixated word per trial, word = presented word per CRA item, 2T = second target word; 3T= third target word. IRR = Incidence Rate Ratio, CI = 95% confidence intervall, p = p-value.

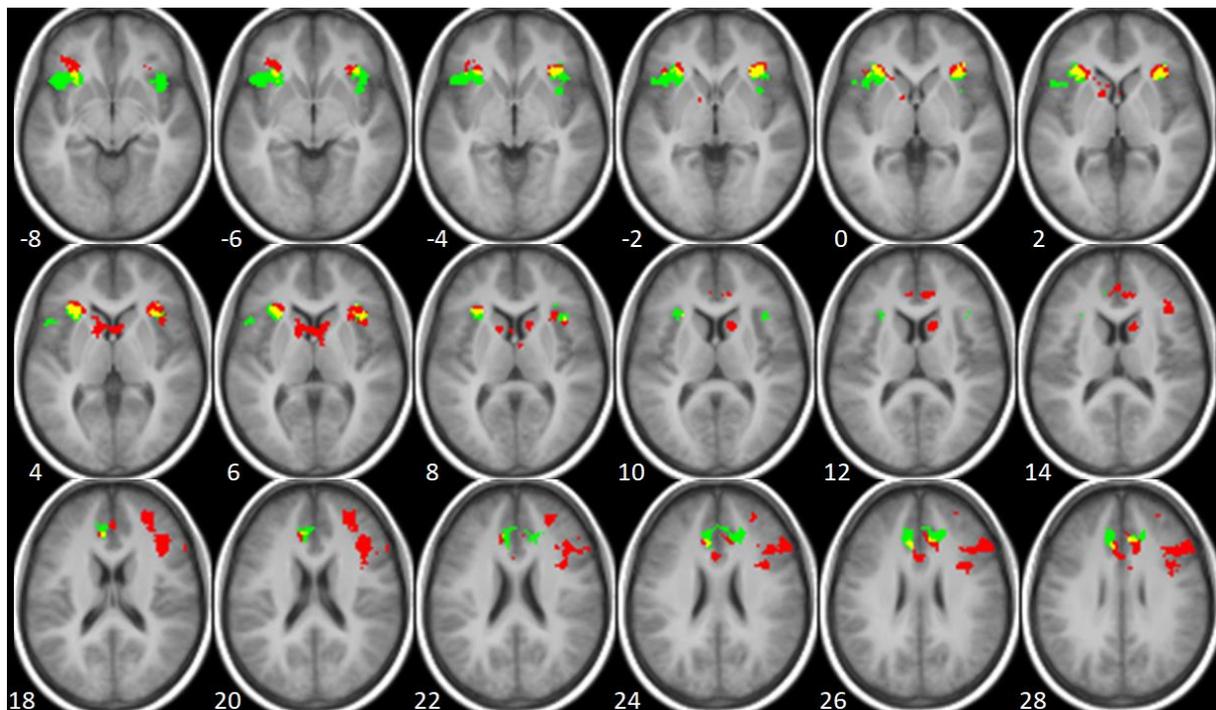


Figure 7. fMRI results for the contrast Restructuring_AHA > noRestructuring_AHA during solution (green) overlaid with clusters of voxels significantly correlating with solution time for

Restructuring_AHA events during solution (red) (voxel-based threshold $p < .001$; cluster-based FWE corrected threshold $p < .05$). The yellow blobs represent the location where the red and green clusters overlap.